FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Probabilistic Load Flow based on Reactive Programming paradigms

Manuel João Teixeira Furtado

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Professor Doutor Cláudio Monteiro

July 16, 2017

© Manuel Furtado, 2017

Resumo

Com a mudança do paradigma do sistema eléctrico no panorama mundial, cada vez mais se observa uma crescente aposta em técnicas de eficiência energética, quer devido à necessidade ambiental, quer económica. Assim sendo, componentes como o planeamento e operação de redes necessitam de ferramentas que auxiliem os projetistas e gestores nas suas decisões.

Durante os últimos anos têm sido avançadas novas teorias que recorrem a métodos probabilísticos, por forma a atribuir pesos a diferentes cenários possíveis de exploração. Estas teorias permitem que ainda se considerem casos muito pouco plausíveis, mas que estes não tenham uma influência muito pesada na decisão a tomar.

Assim, como forma de auxílio às práticas de planeamento e operação, foi desenvolvida a ferramenta *DenDEx* ao longo desta dissertação, que permite avaliar possíveis situações, a partir de um conjunto de dados históricos. Desta forma, assumindo uma natureza probabilística, o método em questão revela-se particularmente útil em situações de ausência de conhecimento prévio sobre as variáveis e também em aplicações em tempo real.

A ferramenta foi aplicada a uma rede elétrica real, com casos reais, por forma a testar a sua validade, eficiência e capacidade de utilização em situações não-ideais. A ferramenta consiste na determinação das variáveis de saída num trânsito de potências DC, como as tensões dos barramentos, com base probabilística, permitindo assim serem introduzidas nas variáveis de entrada diversos tipos de informação, tais como em tempo real ou de previsão.

Em suma, foram feitos testes para diferentes entradas por forma a identificar quais os seus efeitos nos resultados de saída. Foram testadas então quatro situações diferentes: um caso base, com um dado intervalo para as variáveis de input, passando depois para um caso com um intervalo mais apertado. Numa segunda fase, foram introduzidas novas variáveis na forma de potências reativas e manteve-se o intervalo constante, igual ao caso base. Por último utilizou-se, para além das potências reativas, a tensão num dado barramento. O objetivo foi avaliar o impacto dos variados elementos nos resultados de saída

Abstract

With the shifting paradigms of the world wide electric system, each and every day one can witness the growth of energy efficient solutions being applied, due to economic and environmental necessities. Therefore, planning and operation of said electric networks are in need of tools to aid engineers and grid assets' managers in their decisions.

During the past decade new theories running probabilistic approaches have been emerging, in order to assign weights to different scenarios of exploration. This theories still allow the use of a large database, which is consistently built around hypothetical cases, which can be prejudicial if those cases have a very low chance of happening, marking its influence in the final decisions.

As a means to aid planning and operation, *DenDEx* was developed on the course of this dissertation. This new tool allows the evaluation of possible scenarios contained in a matrix of historical data. Since it assumes a probabilistic nature, the method developed is particularly useful in situations such as lack of previous knowledge about input variables and real time applications.

The tool was applied to a real electric system, with real case scenarios, in order to be tested its validity, efficiency and capacity to perform in non-ideal scenarios. The method aims to determine output variables of DC power flow problems, such as bus voltages, with a probabilistic approach, allowing the use of several types of information, either real time or forecast.

In short, tests were done to different inputs in order to identify the effects on output results. It was tested four different situations: a base case, with a given interval for the input variables, and then another case by tightening up that interval. Next, it was inserted new variables in the form of reactive power and it was maintained the interval as constant, equal to the base case. Finally, besides using reactive power, it was inserted another input in the form of a voltage. The objective was to evaluate the impact of the various elements on the output results.

Acknowledgments

I would like to start by thanking my dear parents and brother. Thank you, from the bottom of my heart, for all the support provided during these long six years of my life. Thank you for your education and for setting the best example I could have had during the years. Trusting in me during my very rocky start was fundamental for my success and for that, as well as many other things, I will be forever grateful.

To my grandparents Maria do Céu and Joaquim. Being a second parent is not an easy task, I thank you for the inspiring words and for being the nicest people on Earth. I hope you are proud of my achievement. To my late grandparents Cininha and Manuel, who I am sure that would be proud as well. *Aos meus avós Maria do Céu e Joaquim. Serem uns segundos pais não é tarefa fácil. Agradeço as palavras inspiradoras e por serem as pessoas mais queridas à face da Terra. Espero ter-vos deixado orgulhoso com esta conquista. Aos meus falecidos avós Cininha e Manuel, que tenho a certeza que também estariam orgulhosos.*

To my beloved Tânia. There are little to no words to describe what you have done for me and how much you helped me overcome. My best friend, we have come a long way now, haven't we? Looking back, thank you for being the kindest person, always greeting me with a smile and a hug, no matter what. Also, thank you for being the one that always brings out the very best in me and for keeping me on the track for success. You made me build the courage to face everything, everyday, during the last four years. For everything that you mean to me and for what you've done, I am eternally grateful.

To professor Cláudio Monteiro. One of the very best human beings with whom I fortunately had the pleasure to learn from as a student and to work with as supervisor. An excellent professional, who tries to exceed and improve himself every time. I would like to thank for the availability to meet, week after week, and also for the great guidance and valuable inputs on this dissertation.

Last, but not least, to my friends. A special thanks to Ariana, who has all my respect. We laughed, we cried, we got serious work done and I sincerely regret nothing. One of the best gifts college greeted me with. Thank you hippie! Another special thanks to my fellow countrymen, *"os poveiros"*, thank you all for the moments we enjoyed together. I will never forget my roots and where it all started. To all my closest friends not mentioned above, rest assured: you are not less important. There is a special place for everyone of you in my heart, from my college colleagues to the most distant ones. I can honestly say I have, by far, the best friends in the world, thank you for all your availability and true friendship, I know that I can rely on each one of you.

A huge thank you to all,

Manuel Furtado

"Success is not final, failure is not fatal: it is the courage to continue that counts"

Winston Churchill

Contents

List of Figures

List of Tables

Abbreviations and Symbols

Abbreviations

CDF	Cumulative Density Function
DenDEx	Density Data Extraction
DLF	Deterministic Load Flow
IED	Inteligent Electronic Device
IoT	Internet of Things
KDE	Kernel Density Estimation
LF	Load-Flow
MAPE	Mean Absolute Percentage Error
MCS	Monte Carlo Simulation
MCM	Monte Carlo Method
PDF	Probability Density Function
PLF	Probabilistic Load Flow
PV	Photovoltaic
p.u.	Per Unit
RI	Reliability Index
VM	Variable Modeling

Symbols

P_g	Active Power Generation
Q_g	Reactive Power Generation
P_l	Active Power Load
Q_l	Reactive Power Load
P_{ik}	Active Power Flow through line i-k
Q_{ik}	Reactive Power Flow through line i-k
Y_{ik}	Line Admittance
G_{ik}	Line Conductance
B_{ik}	Line Susceptance
V_i	Voltage on bus i
t_{ik}	Transformer Tap Ratio
δ_{ik}	Phase angle difference between bus i and bus k
α	Shape parameter of Beta PDF
β	Shape parameter of Beta PDF
$h_{p,d}$	Value of the variable d and historical case p
$i(h_{p,d};\alpha,\beta)$	Beta PDF probability of $h_{p,d}$ with shape parameters α and β
max_d	Output maximum value of dimension d
min_d	Output minimum value of dimension d

i'_d	Normalized inputs between min_d and max_d of the dimension d
A_p	Joint activation vector
$h'_{p,d}$	Normalized Value of the variable d and historical case p between max_d and min_d
$maxo_d$	Output maximum value of dimension d
$mino_d$	Output minimum value of dimension d
$h''_{p,d}$	Re-normalized Value of the variable d and historical case p between $maxo_d$ and $mino_d$
I''	Normalized inputs which were selected to calculate the expected value of dimension d normalized output
O''_d	Dimension d normalized output for which is being calculated the expected value
$E[O''_d\|I'']$	Expected Value of a normalized Output O" of variable d with respect to a given normalized input I"
$V[O''_d\|I'']$	Variance of a normalized Output O" of variable d with respect to a given normalized input I"
αo_d	Output shape parameter of Beta PDF
βo_d	Output shape parameter of Beta PDF
A_{top_p}	Percentile of the highest activation values of A_p
NMP	Very small number
f_{obs_i}	Observed frequency
f_{target_i}	Target frequency
NQ	Number of Quantiles
T_i	Target Value
O_i	Output value correspondant to quantile Q_{50}

Chapter 1

Introduction

This chapter has the main purpose to inform about the motivation to write this dissertation and the problem that is proposed to be solved. In order to state that clearly, there is a simple explanation of the document organization and the proposed objectives to accomplish. Also, there will be information about the software used to compute the methods.

1.1 Problem Statement and Motivation

The shifting paradigms of electrical power systems force innovation on said systems due to continuous need of improvement. One of the main problematics are the traditional structures of distribution and transmission power systems, which were built assuming an unidirectional power flow, from upper to lower levels, more particularly from generation points at the beginning of the network to consumption points at the end of the network.

Nevertheless, this approach brings some advantages such as simplicity of operation and management at distribution levels, being the transmissions systems able to transport bulk power over large distances with negligible amounts of losses and the distribution system being able to transform themselves into numerous different operation topologies.

This paradigm did not preview neither automation, nor intelligence, once that the main studies, which were made only to evaluate the security of operation of said grids, were power flow studies related to steady state scenarios. However, new needs in terms of system reliability, distributed generation through renewable energy sources and autonomous control caused a movement to endow the grid with intelligent, controllable and reactive components which manage the grid assets correctly and in real-time.

Along with these new challenges came the need of new power flow algorithms that can cross information such as historic generation/load values, real-time and average power values, output variables of state-estimation algorithms, markets behavior and predictive analysis of generation and consumption into one single method that not only calculates the dynamic power flow over time, but also runs new state-estimation algorithms in real-time.

This can only be achieved through the implementation of intelligent components capable of transmitting information to a central unit, or even between them, in order to have the maximum amount of useful data, which will result in better precision and accuracy of said method.

Currently in the midst of computer age, these new components start to emerge, fully endowed with autonomous functions and communication modules, usually dubbed as smart devices, which allow the implementation of said strategies. The introduction in large scale of those components, often in systems in need of higher levels of intelligence, is a phenomena commonly referred as Internet of Things (IoT) and is probably the key element to unlock the full capability of these networks. [1]

However, this raises some concerns about the systems' security, once the data is not bound to a single entity, but to different administrative domains, managed in a non-centralized way. Reference [1] states the majority of the information in IoT platforms will be stored in cloud and will use cloud computing techniques due to the high scalability of resources.

This alternative inherits the security bound to cloud services, which cannot be attack-proof, putting in jeopardy private information of various users and systems. Thus, it is strictly necessary to have tight security methods to protect such data and also to prevent false information to be fed to algorithms, which can lead into catastrophic events, such as over-voltages, line outages or even blackouts.

In order to link both power flow and smart devices, as it is impossible to predict and control the order of arrival of information from the smart devices about the conditions in which the grids are being exploited, a solution that can suit the problematic at study is through reactive programming. This new programming paradigm might possibly solve the problem, once that it can manage automatically the flow of arrival of information, reacting to external events autonomously and without concerns about the event order or computation dependencies.

One way of observing the subject at hand is through a standard spreadsheet: most of them are created with the premise that it will be introduced constant values into certain cells, called input variables, which will be used in a process of determination of other variables through static formulas. A sudden change in any input variable has an immediate impact in the output variable. That is the basic concept of reactive programming, which is built around the concept of propagation of change in event-driven systems. [2]

This dissertation will approach a new alternative to deterministic load flow problems, linking reactive programming paradigms with a Probabilistic Load Flow (PLF) method. This new alternative may prove that it is possible to automate the electric system, without compromising its operation and exploration.

1.2 Objectives

This thesis is planned to give some insight on the problematics at study, as well as to reformulate and improve the operation of power systems. The main objective is to develop new tools to solve PLF problems, being those tools based on reactive programming paradigms.

The specific objectives contemplate the following topics:

- Evaluation of alternative methods of probabilistic load flow, based on reactive programming concepts;
- Development of data reaction, breakdown and synchronism algorithms;
- Develop a probabilistic power flow tool capable of using different kinds of information;
- Test developed tools on a real case example.

1.3 Document Structure

The present dissertation is divided in 6 main chapters.

Chapter 1 is a brief introduction to the developed work, containing a description of the theme, the motivation to write the present dissertation and the objectives to be completed. It is also described the tools and information used in the course of this dissertation.

Chapter 2 approaches several concepts, background and related work to establish the core contents associated to Probabilistic Load Flow, Internet of Things and Reactive Programming, in order to be able to develop and discuss the present dissertation.

Chapter 3 present the developed methodology to solve the problem stated using probabilistic approaches. It refers to an explanatory case to better understand the results shown in chapter 5, where it is evaluated several indexes such as the reliability index, mean absolute percentage error and average difference between quantiles Q_{10} and Q_{90}. It also contemplates a section on idealized alterations to the methodology to fulfill the interaction between different agents.

Chapter 4 contemplates a brief explanation of the data handled by EDP, about the grid's assets and variables. It is shown how the information was treated and synchronized, the nature of the available data and it is also made a brief characterization of both training and test set.

Chapter 5 shows the results and discusses the values obtained via application of chapter's 3 methodology. Having a total of four different cases tested the objective is to evaluate the behaviour and influence of several aspects of the methodology on the final results. It will be studied the effect of changing the input aperture and the introduction of reactive power and voltage input variables.

Chapter 6 presents some conclusions taken during the writing of this dissertation, analyzing as well the satisfaction of proposed objectives.

1.4 Tools and Information Description

On the course of this dissertation several *softwares* were needed, such as *Matlab* to organize the available data automatically and VBA macros in *Excel*, software in which was implemented the developped tool. The data was kindly provided by EDP Distribuição.

Chapter 2

Concepts, Background and Related Work

2.1 Load Flow

This chapter will address the concept of load flow problems, as well as a brief comparison between deterministic and probabilistic approaches.

2.1.1 Deterministic Load Flow

The main purposes of power flow studies is to aid the planning and designing expansion or creation of electric power systems, as well as defining the best operation strategies of existing systems. Thus, these studies are of great importance and responsibility, once it is the security of operation and maintenance that are at stake.

As investigated before, power flow studies require a set of electrical information in order to calculate the voltage magnitude and phase angle, as well as power (active and reactive) injection and flowing through the lines. [3]

There are several methods to solve deterministic load flows (DLF), namely, through Newton-Raphson or Gauss-Seidel methods, being those methods supported by two Load-Flow (LF) equations, denoting power injection, as follows:

$$P_i = V_i \sum_{k=1}^{n} V_k (G_{ik} cos(\delta_{ik}) + B_{ik} sin(\delta_{ik})) \tag{2.1}$$

$$Q_i = V_i \sum_{k=1}^{n} V_k (G_{ik} sin(\delta_{ik}) + B_{ik} cos(\delta_{ik})) \tag{2.2}$$

As usual, there are three types of buses in a power system: PV buses with specific active power injection and voltage magnitude, PQ buses with specific active and reactive power injection and the REF (often named Slack Bus) with specified voltage magnitude and phase angle.

To solve deterministic problems, generally the method starts with the assumption of phase equilibrium and with specific values for voltage magnitude and phase angles, often called flat start, being given a constant value to the REF bus. From there on, it is assumed constant grid parameters and the method is then applied.

From the solution of said methods it is possible to determine the power flow from bus "i" to bus "k", denoting line ik, through the following equations:

$$P_{ik} = -t_{ik}G_{ik}V_i^2 + V_iV_k(G_{ik}cos(\delta_{ik}) + B_{ik}sin(\delta_{ik})) \tag{2.3}$$

$$Q_{ik} = t_{ik}B_{ik}V_i^2 - B_{ik}^{'}V_i^2 + V_iV_k(G_{ik}sin(\delta_{ik}) + B_{ik}cos(\delta_{ik})) \tag{2.4}$$

However, for current electrical systems, these DLF studies do not reveal the true scenario of operation, being the output of said studies a snapshot of the instant mode of operation, when the data was collected. That may not aid in terms of reliability once that there is some variation related to load regimen, operation strategy adopted or even grid topology, that is, deterministic criteria often ignore the variability of important variables and parameters, affecting the design and operation policies of power systems. Also, this studies are often based on "worst-case scenarios" which may not include normal state of operation and sometimes even scenarios that are highly unlikely. [4]

2.1.2 Probabilistic Load Flow

Contrarily to DLF, PLF studies aim to account uncertainty in the methods, in order to have a more accurate solution. This problem dates back to 1975, when many researchers have equated the most efficient and accurate method to obtain PLF solutions: state vector and line flows. [5] [6] [7] [8]

Reference [5] defines the numerical techniques developed to solve load flow problems instead of using deterministic methods with the main objective of reducing computational time and storage requirements, while conserving high accuracy levels.

Reference [7] addresses to the use of least square estimation techniques through linearized models, being also discussed a second order model and a iterative computation procedure which aims to compare both techniques.

Reference [6] utilizes a Monte Carlo Simulation (MCS) to address to the effects of nonlinear network equations and the assumption of normal distribution for output variables. It is also presented an algorithm to overcome the nonlinear effects and give new impetus to PLF analysis.

Reference [8], similarly to [6], also addresses to nonlinear network equations and dispatching strategies. The technique is applied to second order approximations for the load flow and current equations, through a noniterative computation model of means and standard deviation for the several outputs. Also, generation outages are separately treated in the course of the previously referenced paper.

2.1.2.1 Information

The input variables provided by electric networks are not usually the best data to analyze, since there are problems associated to them, such as modularity, aggregation or abundance. The information is not usually referring to the same instant in time, as measures scheduled individually and not coordinated with other agents. Associated to this problems, there is a problem as well about the abundance of measures, with those being available only in certain key spots of the grid (modularity).

Firstly, to apply the previous methodologies, there is a need for an abundance of historical data, which sometimes can be difficult, with special emphasis on distribution systems, due to the lack of measurement devices.

Secondly, information is often held by different owners and that fact may hamper its management. A strategy to overcome such difficulties may be through the creation of new business models, operating such information as using cloud techniques, in order to centralize all information in one place. The current limited access to such data is still one important barrier of the development of PLF methodologies.

Thirdly, intelligent electronic devices (IED) receive all the information disaggregated, either in time or type, this is, IED's microprocessors acquire the data from sensors, such as voltages, currents or power, but they lack synchronism, as some are instant and others are averages. Often, measurements signaled as synchronized correspond to different points in time.

Lastly, there is a tendency to include predictive information calculated separately, in order to complement the historical information and advise the distribution system operator of future conditions and operation states.

Given the stated above, researchers such as [9] or [10] address to the techniques used to build the input variables, which will be discussed later on in this dissertation. One important aspect is that electrical variables, as P_g, Q_g, P_l, Q_l and Y, which represent power generation (active and reactive), loads (active and reactive) and network configuration, respectively, are best defined through their joint density function, assuming dependency between them.

Relatively to network configurations, Y should be a discrete three dimensional matrix, denoting conductance, susceptance and probability of occurrence of each configuration state, since it can change instantaneously and can not be simultaneously in two different states. [4] It also could be added a fourth dimension denoting time spent in each state.

In order to reveal the dependency between variables, Pearson coefficients shall be evaluated. Pearson coefficients, usually represented by ρ, is a measure which indicates the relation between two variables, X and Y, in the interval $[-1;1]$, where 1 is total positive linear correlation and -1 is total negative linear correlation, and can be calculated as a function of covariance between the two variables, and their standard deviation, as follows:[11]

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.5}$$

The figures below shows some examples of Pearson coefficient values between two random variables.

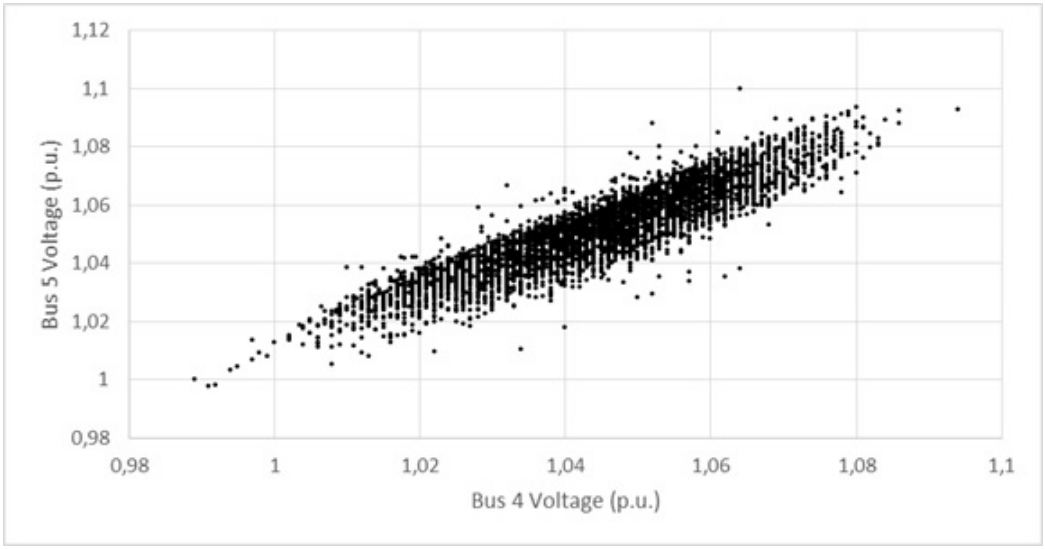

Figure 2.1: Example of spacial correlation between nearby buses voltage on a real case

As it is evidenced above in figure 2.1, assuming bus 4 and 5 are connected, there is an obvious and strong spacial correlation between the voltage of two nearby voltages. Thus, one should not assume independence between interconnected buses, which simplifications assuming independence could carry gross errors into certain models.

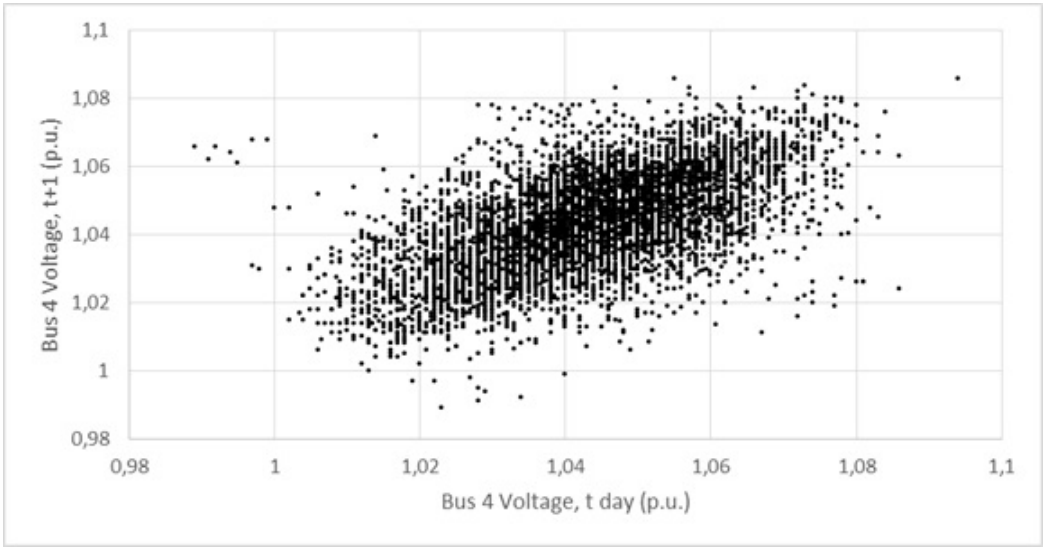

Figure 2.2: Example of time correlation between voltages on a bus, instant t and $t+1$

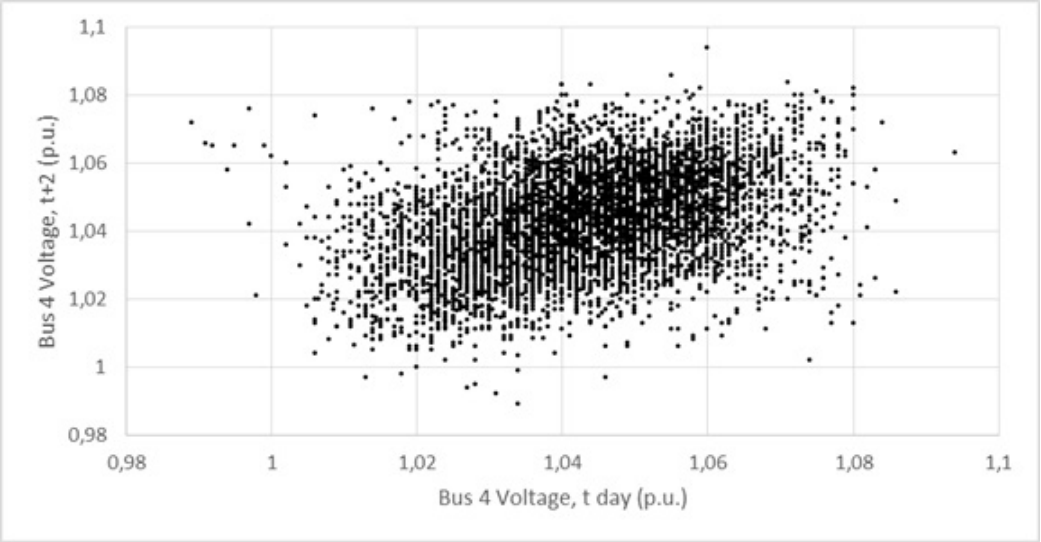

Figure 2.3: Example of time correlation between voltages on a bus, instant t and $t+2$

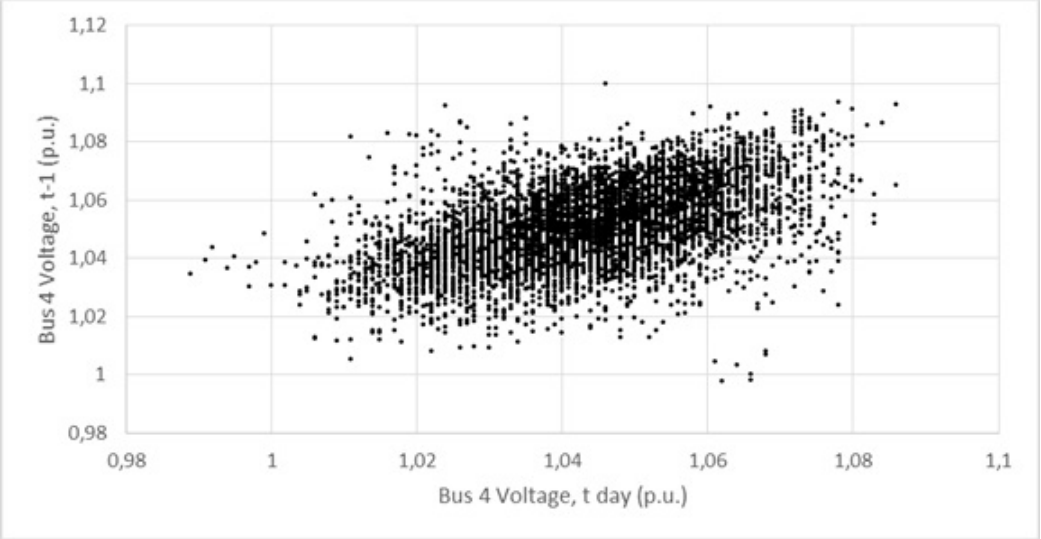

Figure 2.4: Example of time correlation between voltages on a bus, instant t and $t-1$

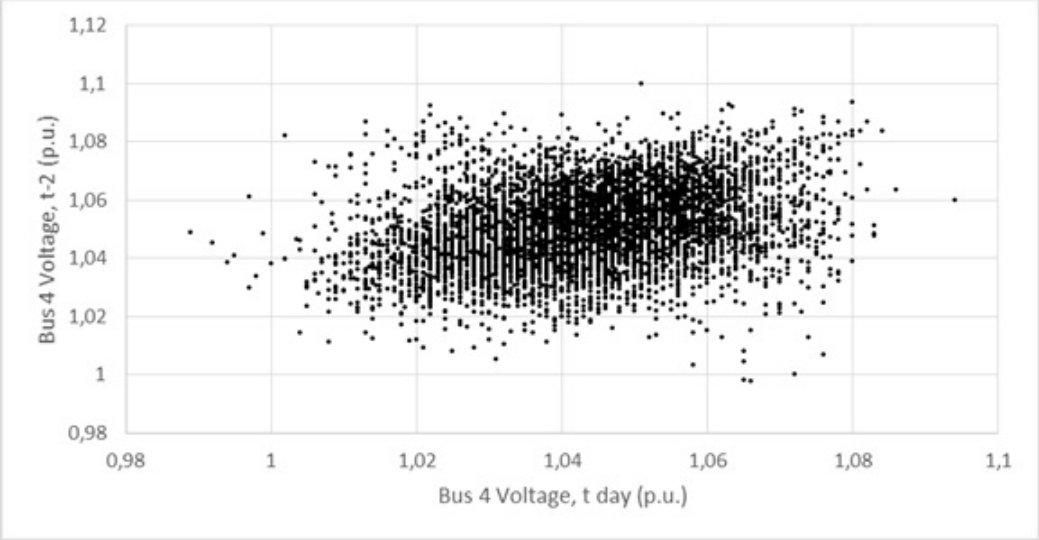

Figure 2.5: Example of time correlation between voltages on a bus, instant t and $t-2$

As for time correlation, one can observe the lack of intraday correlation, which gets significantly worse as the chosen days go further back in time, this is, variables are less correlated the more separated in time they are.

Another interesting and important point is the evaluation of the multiple sources of uncertainty. As stated before, the main objective of PLF studies, relative to DLF studies, is to account such uncertainties, so the output of said studies are a more accurate representation of the system.

Data collection is generally done in substations through the use of transformers (current or voltage), which is then sent to the system operator. Given this, as a correct representation of input variables is needed and measurement errors are not a rare event, there is a need to assess the consistency in order not to make important decisions based on false information. The sources of information and its associated uncertainty can be such as follows:

(a) No measurement devices which translates in assuming physic properties of grid components, such as load, generation and lines minimum and maximum values;

(b) Measurement devices, accounting for measurement errors and more properly the device accuracy;

(c) System operator information, which can be compressed in several non-synchronized intervals, of either intra-variable or inter-variable nature;

(d) System operator raw information, as a set of multiple measures in a single interval.

According to [12], uncertainties can grouped by their nature: aleatory or epistemic. Aleatory uncertainty is defined as the variation of given variable characterized under a PDF, constructed using sufficient samples of a stochastic process. Epistemic uncertainty (also called ignorance uncertainty) is the uncertainty derived from the lack of knowledge, either of the surroundings of

the system at study or even experimental data. This kind of uncertainty is usually represented as an intervalued quantity between a given range, where any value is as true as another in that given range.

As different levels of errors can be encountered as a result of previously stated natures and sources of uncertainty, there is also different ways to treat such information, or the lack of it, without compromising the results. The previously presented list is a peculiar form of aggregation in order to distinguish different information level requirements. This differentiation is particularly useful to pair the correct variable modeling (VM) with the available information, as some techniques are consistently used without correctly identifying the minimum information requirements to successfully applied. Some VM techniques are presented as follows:

- Interval, information about range of values $[min; max]$;
- Fuzzy sets, with added information about most plausible point of operation;
- Parametric PDF:
  - variable modeling through subjective parameter values, provided by system operator;
  - parameter estimation from available data.
- Discrete PDF, representing frequency of occurrence of a given variable;
- N-tuple representation of original data.

There are several ways to treat the information used to nourish PLF studies, but the most common way is to convert input variables into PDF, using approximations. Most recent literature identifies the types of distributions used to describe both load and generation, such as [13][14][15][16][17][18][19].

Generally, the vast majority of researchers use normal distributions to describe load behavior [13][14][15][20][21], discrete distributions to describe dispatchable generation [22], Weibull or gamma distributions to describe wind power [10][13][14]. There is still a great uncertainty about photovoltaic (PV) most suitable distributions, hence it is commonly applied a normal distribution as well [23]. The main purpose that drives several into using PDF is to take into account deviations and uncertainties, otherwise ignored with deterministic methods.

Although the use of several different parametric distributions is the most common approach, there are other variable modeling techniques which can add value into other methodologies, such as Fuzzy Sets [21] or Scaling Factors [24].

Fuzzy Sets [21] technique is mainly used when there is insufficient statistic information to assign a probability density function to a given variable and the evaluation of the system variables is performed using fuzzy arithmetic calculations. Reference [21] states that authors commonly modify fuzzy calculations, since fuzzy arithmetic algorithms barely give viable power flow solutions.

Scaling Factors [24] is particularly effective when neural network methodologies are applied since one of the main characteristics of neural networks are sigmoid activation functions of the

neurons. Sigmoid activation functions perform better when input data is between a certain range, typically between [-1;1]. Defining correctly such techniques are of great importance once that it is easier to cover a wide range of load and generation levels, their variations and, at the same time, obtain better results. Such techniques were mainly applied to small sized power systems, yet to be applied in real-scale power systems to prove the effectiveness of such methods.

The following table synthesizes all of the stated above.

Table 2.1: Variable Modeling alternatives applied in probabilistic load flow studies

Information / **Variable Modeling**	**(a)**	**(b)**	**(c)**	**(d)**
Interval	[24]			
Fuzzy	[21]			
Subjective Parametric PDF	[14][15][17][18]			
Statistical Parametric PDF			[10][20][23]	[13][16][19]
Discrete PDF				[22]
N-tuple				[25]

2.1.2.2 Approaches

Throughout this section methodologies are described and separated under some aspects considered relevant, such as methodology itself and LF equation use. In [22], PLF methods are just separated into two big groups: numerical and analytical. This section opts for the separation into smaller groups as it will be discussed further.

Numerical approaches use Monte Carlo Simulations (MCS), performing DLF algorithms to a large number of random data, corresponding to, ideally, infinite cases of injected power and voltages. The great precision of such methods, due to the large number of cases studied, makes it the method adopted for comparison with others, namely analytical methods, or combinations of both.

Analytical approaches, in their very essence, use arithmetic calculation, such as convolution techniques, or a combination between convolution and MCS approaches. Convolution based, and analytical methods in general, tend to have some problems related to the non-linear nature of LF equations and also the correlation between variables, which are not frequently independent or linearly correlated.

In order to solve PLF studies with analytical methods, there is a need to assume some approximations:

- Linearization of LF equations;
- Full independency or linear correlation between variables;
- Parametric distributions (normal, gamma, Weibull,...) modeling variables such as load or generation;

- Constant network data (configuration, parameters,...)

Given this, the present section focuses on separating the latest published methods, in order to do an overview of the present status of PLF studies.

Table 2.2: Probabilistic Load Flow Methodologies

Methodologies	**Reference**
Monte Carlo Simulation	[6][22][25]
Fuzzy Sets	[21]
Point Estimate Method	[20][26][27]
Cumulant	[28][29][19]
Latin Hypercube Sampling	[30]
Neural Networks	[24][31]
Unscented Transformation	[15]

Table 2.3: Load Flow Equations Treatment in load flow problems

LF Equations Treatment	**Reference**
No treatment	[22]
Krawczyk Method	[32]
Cornish-Fisher	[28]
Fresnel Triangle	[33]
McLaren's Series	[27]
Gram-Charlier Expansion	[29]
Taylor Series Expansion	[24]
Multiple Linearization Points	[25]
Forward/Backward Sweeps Method	[21]
Nataf Transformation	[20][30]

The previous tables organize the information gathered throughout the research made on this topic. Due to the large numbers of papers and combinations of methodologies, it was previously chosen to treat each point separately, variable modeling and LF equation treatment respectively. A brief explanation of each will be given further in this section. Algorithms presented in the following references are not going to be fully explained, as it is not the purpose of this section.

Firstly, there are several methodologies that can be applied to solve PLF problems. The great majority of them are analytical methodologies, which were created to simplify the calculation in order to obtain accurate values with the least possible computation effort, with exception to MCS methods, numerical methods which constitute the comparison standard, due to the method's great performance.

Given this, MCS methods, described in [22] propose the utilization of a random number generator to identify all possible case of power generation and loads in order to solve multiple DLF problems, using non-linear LF equations. After obtaining multiple DLF solutions, said solutions are then stored in a vector and organized by frequency of appearance, constructing the PDF for a given random variable.

In [21] fuzzy sets are used to solve the PLF problem. The theory behind fuzzy sets is being recognized as one of the tools with most potential to solve said problems, once that fuzzy membership functions establish relations between possibilistic variables and their degree of uncertainty. The method also enables "Human Judgement", being possible to define a range of values to uncertainty.

Reference [20] used a confined number of estimating points with different weighting coefficients, whom could be obtained through Gaussian-Hermite integration. The method proposed the calculation statistical moments of output random vector using the first three points of random variables joint probability density functions to replace the joint density functions themselves. The algorithm in [20] is a variation of the method proposed by Hong described in [26], which can be found respectively in each reference.

In [28] and [29], it is proposed the use of probability distributions defined through cumulant-generating functions as an alternative to the moments, helping the simplification of theoretical problems. The main difference between them is the methodology used to construct the Cumulative Density Functions (CDF), using Cornish-Fisher expansion and Gram-Charlier, respectively. Others variable modeling methodologies are adopted with Cumulant methods, as represented in [19].

In [30] is stated that Latin Hypercube Sampling method is, in its essence, a combination between Monte Carlo Stochastic Simulation and a Quasi-Monte Carlo Simulation, solving the PLF through correlation of input variables that follow arbitrary distributions.

In [24], besides highlighting the lack of attempts to solve PLF using neural networks, also stated in [31], the author also reveals a problem of using artificial networks, which include the use of a great number of neural networks. This constitutes a problem itself, as it turns out to be rather unattractive and computationally heavy. Although, [24] separated the analysis into two methods, the first using Backpropagation Algorithm and the second also trained with Quickprop Algorithm. The first method relies on the steepest descent technique to minimize errors and reports directly the confidence levels of the stochastic output quantities. The second method uses second order derivatives of error functions, which in turn accelerates learning rate.

Lastly, in this organization model it is approached the techniques to deal with the non-linearity of LF equations. As mentioned before, LF equations are not linear, which is imperative to easily perform PLF studies using analytical approaches. [22] As the vast majority of methods are analytical, with exception to MCS, some linearization models are needed. In MCS-based methods, there is no need to linearize LF equations since the method uses a vector of random variables and perform multiple DLF methods iteratively. Given this, there are several techniques to linearize, without compromising results, LF equations, some of them exemplified on table 2.3, being the most used the Cornish-Fisher, Gram-Charlier and Taylor Series expansions.

According to many researches such as [34], Cornish-Fisher expansion produce significant errors in tail parts, this is, the tip of the distribution curves, while Gram-Charlier is only suitable when applying Gaussian distributions. Due to the limitations of parametric estimation methods,

namely with respect to renewable energy sources, there has been a shift to nonparametric estimation, such as Kernel Density Estimation (KDE) methods, due to the difficulties of KDE on dealing with correlativity.

Essentially, KDE is an inference tool, based on finite sample data of a given population. In order to approximate mathematically a given function, there is two main components of estimation: $KernelK(\bullet)$ and a smoothing component h known as $Bandwidth$, both applied in the following equation, where $\hat{f}_n$ denotes the kernel density estimate of overall density function f_n at a point x. [35]

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{2.6}$$

Reference [30] also states that even KDE tools can not identify PDF precisely, as it performs badly, with considerable errors in marginal intervals with high probability.

2.1.2.3 Applications

This section will briefly discuss the major applications of PLF studies. The increased uncertainty of power system operation and planning led to the use of new alternatives to DLF studies, as it is mandatory to obtain more accurate values of reliability. Hence, the need to use probabilistic approaches is obviously connected to the type of problem operators want to solve with power flow.

If the objective is to evaluate extreme situations, the correct way is to apply PLF methods, since they allow the study of occurrence of a given extreme value associated with the respective probability. As an example, when discussing the probability of exceeding the power flow limits of lines, reverse power flows or disperse generation management, uncertainty should be clearly identified and modeled.

Most PLF problems are not solved to obtain the expected value, but to obtain a given event's probability of occurrence, for example, the probability of an event where the load flow is greater than the rated power of a line. Expected value is usually useful for state estimation algorithms, which could also be obtained via PLF studies, carrying evidently more reliable and useful information than solely the expected value.

Traditionally, extreme scenarios can also be modeled using DLF, if an extreme type of operation is considered. This type of evaluation can be misleading for decision makers, since they may lose track of the probability associated to a certain extreme scenario, which often leads to wrong and extreme decisions, without necessity to do so.

In short, the previous paragraph presents the justification why most studies of planning and risk assessment should be probabilistic and not deterministic. Extreme scenarios should be formulated within the scope of pretended applications, since it is possible to evaluate scenarios for different time horizon and intervals, giving them greater importance to correctly fit in the type of application.

Intervals could be studied in terms of an hour, one year or ten years, for example. Obviously, smaller intervals means narrower variables' variances which in turn correspond to different and more defined uncertainties.

Time horizon defines the future perspectives from the planning point of view. A long range time horizon is often associated with greater uncertainties in planning, unlike short range time horizons, which are commonly associated with information uncertainties in operation.

Essentially there are two main groups of application of PLF studies to planning: long term and operation. In order to correctly handle the information, variable modeling should be divided into two categories, as stated in [36]: short term and long term modeling. Short term modeling is applied to system operation and, on the other hand, long term modeling is applied to system planning. [22]

Data collected in the systems' substations contains, amongst other information, for short term data, daily peak values for the random variables, generally for the period of two months and, for long term data, annual peak values for a given number of years. PLF studies are then applied with the methodologies studied previously in section 2.1.2.2.

Long term planning, often referred as grid expansion planning, uses long range horizon information, which can reach ten, twenty or even forty years if the component being studied is a substation, for example. To study this situations, uncertainties are commonly modeled based on load growth forecast in existing buses and also based on the location of new substations forecast, granting this LF problems new components.

Besides, grid expansion planning also studies extreme operation states, having uncertainty in extreme scenarios, as well as uncertainty related to the synchronism/maximum load correlation due to several aspects as different seasons or demography fluctuations. Within extreme operation states, there is one special kind of extreme state which is a new paradigm for electric systems, consisting in the inclusion of disperse generation. However, the future power output of a certain bus, or production of a given power plant, is not in the aim of said studies. Instead, this studies account for the future installed capacity and location of power plants.

Short term planning, often referred as operation planning, has as main objective the study of extreme scenarios with the present installed capacity. In this cases, time horizons are commonly of one year and are studied all possible cases that may occur in the said year, either being likely or unlikely, including all periods of the day. When studying short term planning, it is usual to vary the intervals at study between daily, hourly or even quarter of an hour intervals, depending on the level of detail needed. This kind of planning allows the decision maker to analyze the grids' financial effectiveness, active power losses and also to calculate the probability of disperse generation management needs.

It is important to point out that the type of information, discussed previously, depends on time horizons, for example, if one year is selected as time horizon, there is an uncertainty within the operation for that year; if one week is selected as time horizon, it is possible to include details of disperse generation, such as wind or solar; if one day is selected as time horizon, the uncertainty associated to the load values differ themselves within the time of the day. As stated before,

by shrinking the time horizons, typically under one hour, uncertainties tend to be restricted to measurement errors and lack of synchronism within the collected data.

This topic has an important role in the use of PLF methods and the research made pointed out that most of recent papers studying methodologies just describe the method itself and does not state neither general applications of PLF studies, nor the best application of the method at study. Also, this topic lacks updated research, since the most useful papers found about the topic are dated from 1976 to 1991.[36][37][38]

2.2 Internet of Things

The concept was born in 1990 as John Romkey developed a toaster connected over the internet, using a TCP/IP networking, whose project consisted essentially in simply turning on and off the said toaster. Although, the first use of the IoT idea belongs to Kevin Ashton, in 1999, working at Procter&Gamble (P&G) at the time, when he used the term to describe, in a presentation to the executives, the integration of Radio-Frequency Identification (RFID) in P&G supply chain. [39]

A common definition, as stated in [40], is "represents the convergence of a variety of computing and connectivity trends that have been evolving for many decades."

European Research Cluster on the Internet of Things (IERC) [41] also defined IoT as "A dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network."

Given this, there is a special interest for system operators in the integration of such technologies in industry or power systems. General Electrics [42] considered, in a recent publication, that the convergence between advanced computing and the global industrial system would profoundly transform many aspects of daily life.

Such concept is already in application in smart grids, which assumes different components of a given grid should be connected and possess sufficient intelligence to operate autonomously. The actual challenge is to add new functionalities to the components such as the ability to run power flows and opt for different network configurations automatically, in order to always operate under the optimum state.

In the company point of view, there are certain needs in term of operation in which IoT-based solutions can really make an impact. The needs are such as follows:

- Improved productivity – directly affects the success and profits of companies;
- Market valorization – different solutions and products can add market value to companies, if successful;
- Cost effectiveness – one of the main drivers of modern companies is the reduction of operational costs, which is one of the main purposes of using intelligent devices and solutions.

In power systems utilities, the application of IoT concepts is in motion, for example, in projects such as *InovGrid* in Évora, by EDP Distribuição, which includes the integration of IED's to manage systems' assets. This project is proving to be one of the most revolutionary ones in the distribution grids, since it predicts an evolution towards a high number of small producers within the distribution network (lower levels of voltage) and reconfiguration in case of any component outage.

However, this technology raises some questions, namely about where to store all the information to be used by these grid assets. One technology that seems consistent and mature enough is the cloud environment, since it allows the sensors not to have coupled storage. Instead, sensors send the information directly into an online database, ready to be used by other components.

Given this, some properties of IoT systems are of utter importance to the thematic at study and, more properly, this dissertation, such as self-configuration, self-protection, self-description and self-energy-supplying.

Self-configuration allows a quick configuration of the system, which otherwise would take longer if done by hand. Besides providing remote configuration applications, a system with the possibility of self-configuration, allied with intelligence to determine the optimum operation point, should also autonomously reconfigure the systems' topology.

Self-protection is defined as autonomously changing security and privacy levels, in order to avoid attacks with the aim of obtaining private data. This functionality should not affect the quality of service and experience.

Self-description is particularly interesting, as it allows the combination of different devices such as metering systems and sensors, which should be able to describe themselves, their characteristics, capabilities and data, in order to interact with each other.

Self-energy-supplying employs sustainable solutions into intelligent devices, such as renewable energy sources as power supplies or the ability to harvest energy from other sources.

According to [1] there are some other properties of IoT systems that need further research in order to optimize some features: self-adaptation; self-organization; self-optimization; self-healing; self-discovery; self-matchmaking.

Besides properties, other concerns emerge, mostly related to data management, as is constitutes a crucial aspect of IoT applications, considering the fact that Machine-to-machine (M2M) computing is on the rise. Thus, there is a need to store great amounts of generated data by Data Collector and Analysis (DCA) modules, cores of IoT applications, and to facilitate the handling of such information, easily and securely.

Some of the main functions of DCA modules are such as following:

- User data storage – data collected by sensors is then stored in a cloud environment;
- User data and operation modelling – enables the creation of data models to correctly store information received from sensors and prepare it to be used by other users;
- On-demand data access - application programming interface (API) to access easily to data;

This all can not be achieved without working out a key element of IoT applications, which is adequate security. Currently, and in the future, systems will continue to grow, as well as the number of users of said systems. Given this, there is a need to protect the information, since it has great value, as well as it can be dangerous if badly handled. So, in large-scale applications and services that base their use in IoT infrastructures there is a special need to do some research in order to advance in critical areas to shield against malicious intents, such as DoS/DDOS attacks or access from not allowed credentials. Given this, if security does not evolve in the coming years, applications running the IoT concept will resemble more like an Intranet server, drastically different from an Internet one.

Reference [1] predicts on chapter 5.1.2 smart city data, acquired from different modules, will use cloud storage and computer techniques, due to high scalability of resources and reduced cost of operation and maintenance. Thus, smart city applications will inherit cloud computing security and risks, as well as the need of authenticate and authorized access to be provided with trusted information from sensors and actuators.

This kind of structure is exposed to several types of attacks, as discussed previously, that could pose different troubles for the said system, ranging from small interferences to exposure of private information and system outages. Hence, an attacker could possible infiltrate in different layers of the structure, such as manipulation of sensor measurements, obtain credentials from sensors, impersonate an intermediary to withdraw information (phishing) and actuate on systems' assets.

One can observe now that there are several challenges to overcome before applying the previously evidenced technologies on large-scale and vital systems such as electric power systems. The main concerns revolve around security, storage and autonomy of IoT appliances. This concept is identified as a central topic to correctly approach the reading of this dissertation, as it is involved in the main computing characteristics of the model presented.

As discussed previously in this dissertation, there are mainly two types of applications of PLF studies: long term planning and short term planning. From the long term planning point of view, there is little interest in the information supplied by IoT applications, since long term planning uncertainties are derived from hypothetical scenarios and not from information collected via IoT.

However, considering short term planning, there is a growing interest in IoT applications, since uncertainties take in account real time measurements, which are held by different owners, connected to the electric system. Some of them include SCADA applications present in distribution and transport systems, energy markets information, load and self-consumption levels in distribution grids, dispersed generation information, generation connected to industries, such as Combined Heat and Power (CHP) and, finally, information from installed smart meters in different users of distribution systems.

Given this, it is possible to infer that there is plenty of information badly handled by the different owners. One must ponder new paradigms in which information is stored solely in one place, properly controlled by an experienced regulatory body, ready to be handled by the modules of different operators, this is, distribution networks applying information from transmission systems and vice versa. This may result in more accurate power flow algorithms and solutions and,

consequently, better planning of distribution grids.

The current paradigm features software like SCADA, which is too complex to handle information variation and new interactions on real time. Even on distribution networks, the future tends for having disaggregated information on substations, where components communicate between them and in a cloud environment, without the need of SCADA softwares.

For the writing of the dissertation we will assume that the existing information is independent, distributed and accessible, not structured in an organized way, but a traditional one, ready to be used in power flow algorithms. Information is not synchronized, not measured in the same time intervals and can have different measurement units (SI or per unit) or even substitution variables, for example, wind speed rather than wind power output of a given park.

The idealized methodology will be capable of using this information without having to structure it in advance, because the intended approach is itself based on fragmented and distributed algorithms that also do not have synchronous and structured communication between them.

2.3 Reactive Programming

The current subsection aims to present a known paradigm for event-driven systems. Taking distribution energy systems as example, the number of times the system states change is considerable and based this fact it was developed the idea of turning sensors and actuators autonomous using reactive programming in order for them to react in real-time to real-events.

This paradigm is not new, although it was never applied to such a large system. Reference [43] defines reactive programming as part of three types of computer programming: *"It is convenient to distinguish roughly between three kinds of computer programs. Transformational programs compute results from a given set of inputs; typical examples are compilers or numerical computation programs. Interactive programs interact at their own speed with users or with other programs; (...) Reactive programs also maintain a continuous interaction with their environment, but at a speed which is determined by the environment, not the program itself. Interactive programs work at their own pace and mostly deal with communication, while reactive programs only work in respond to external demands and mostly deal with accurate interrupt handling. Real-time programs are usually reactive."*.

The information stated above enables the application of the concept into this dissertation: the primary objective is to create a model that runs solely on information gathered in real-time from near modules. A simple example to explain the use of reactive programming is pretending buses 2, 3 and 4 are connected to bus 1: the idea is to use information from buses 2, 3 and 4 modules and run an autonomous probabilistic load flow algorithm, in order to comprehend the state of operation of the system, endowing the grid with knowledge to change topology to reach the optimum point of operation. A purely reactive method based on probabilistic approaches, to account for uncertainty and slight variations, and IoT concepts to manage grid data and store results for future improvement of the models are the scope of this dissertation.

Chapter 3

Methodology

3.1 Density Data Extraction - DenDEx

The method applied in this thesis will be based in the NW-KDE method, whose principles are based on the work of [44] and [45]. The main difference between the NW-KDE method and a standard KDE method is that, although this method tends to be more complex mathematically, it allows the user to determine the Beta PDF parameters solely from its expected value and variance.

A Beta PDF is defined in a finite interval between a minimum and a maximum, 0 and 1 respectively, and is composed by two positive shape parameters α and β. Essentially, a Beta distribution PDF assumes values of probability as follows:

$$i(h_{p,d};\alpha,\beta) = k \cdot h_{p,d}^{(\alpha-1)} \cdot (1-h_{p,d})^{(\beta-1)} \tag{3.1}$$

In equation 3.1, $h_{p,d} \in [0;1]$ and k is a constant. A brief example of the usage of different values of α and β is presented in figure 3.1.

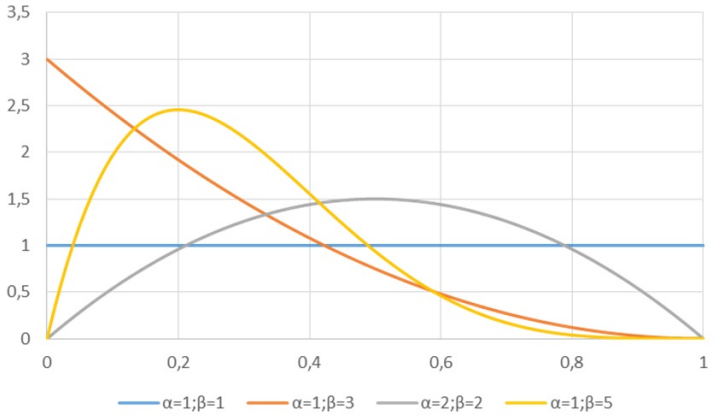

Figure 3.1: Example of Beta PDF

Being this method applied to historical sets, the said expected values and variances of input and output variables can be later translated into useful probabilistic information, either to assess present state of operation or to probabilistically forecast other states of operation, that is, if some information about load levels or generator dispatch is known beforehand. Some changes were done to the original NW-KDE method in order to apply specifically to the problematic at hand and to create a new method to extract data from a knowledge matrix.

Given this, this dissertation considers an historical set of m variables with n historical cases. Therefore, the data collected composes what is called a knowledge base dataset, from now on referred to as knowledge matrix, shown in 3.2.

$$\begin{Bmatrix} h_{1,1} & \cdots & h_{1,d} & \cdots & h_{1,m} \\ \vdots & \ddots & & & \vdots \\ h_{p,1} & \cdots & h_{p,d} & \cdots & h_{p,m} \\ \vdots & & & \ddots & \vdots \\ h_{n,1} & \cdots & h_{n,d} & \cdots & h_{n,m} \end{Bmatrix} \tag{3.2}$$

As an example, one can assume the variables of the knowledge matrix to be referred to an electric system. Each dimension of 3.2, or each column, represents a given electrical variable, such as voltage, and each historic case, or each row, corresponds to a given measure of the said variable.

In order to proceed the application of the method, inputs are needed to be as represented by 3.3. Note that each input is referred to one dimension, as it is necessary to determine the dimensions' maximum and minimum values. Besides that, one can add relevant information, through specifications of α_d and β_d, shape parameters of the dimension d, only if there is an accurate knowledge about the said shape or, in some cases, just by shortening the interval of a dimension by adjusting maximum and minimum.

$$\{i_1(min_1, max_1, \alpha_1, \beta_1); \cdots; i_d(min_d, max_d, \alpha_d, \beta_d); \cdots; i_m(min_m, max_m, \alpha_m, \beta_m)\} \tag{3.3}$$

It is of utter importance to normalize the data being fed to the method since, in this case of application of Beta PDF, it only works in a close interval $\in [0,1]$. One way to normalize data is by using equation 3.4, where h' denotes the normalized data and h denotes the denormalized data.

$$h'_{p,d} = \frac{h_{p,d} - min_d}{max_d - min_d} \tag{3.4}$$

Given this, the inputs presented in 3.3 need to be normalized under 3.4. Therefore, the normalized inputs take another form, presented by 3.5 since it is already assumed that we are working in a close interval $\in [0,1]$.

$$\{i'_1(\alpha_1, \beta_1); \cdots; i'_d(\alpha_d, \beta_d); \cdots; i'_m(\alpha_m, \beta_m)\} \tag{3.5}$$

After retrieving the normalized inputs, it is calculated for each historical case the joint activation value, also known as A_p, through the product of all single activations, for a given p case, from each active input variable, calculated by 3.6.

$$A_p = \prod_{d=1}^{m} I'_d(h'_{p,d}; \alpha i_d, \beta i_d) \tag{3.6}$$

In equation 3.6:

- A_p is the joint activation vector;
- $I'_d(h'_{p,d}; \alpha i_d, \beta i_d)$ is the normalized input containing the values $h_{p,d}$ and the shape parameters αi_d and βi_d of the active variable d;
- $h'_{p,d}$ is the normalized value corresponding to the active variable d and historical case p;
- αi_d and βi_d are the Beta PDF input parameters of a given d dimension.

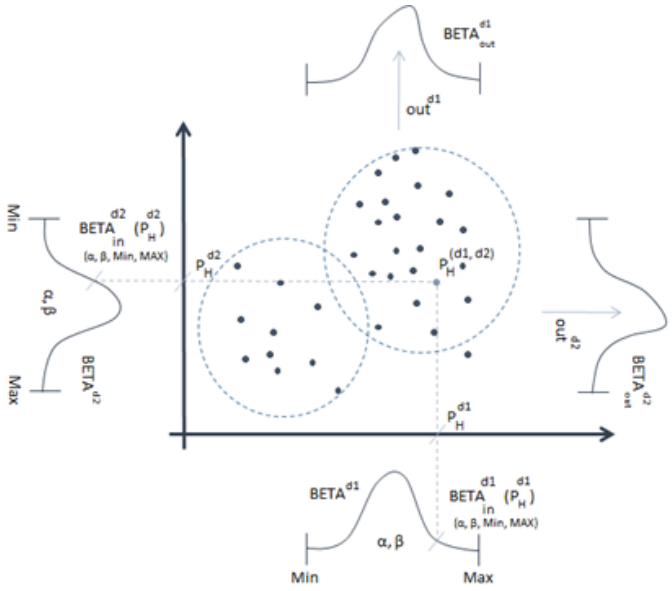

Figure 3.2: Example of Beta PDF

Figure 3.2 gives some graphic insight about the calculation of activation values. This example evidences the calculation involving two variables, denoted by dimensions d_1 and d_2. As one can see, historic case P_H is referred to both variables, having a different activation value for each one, depending on both α and β parameters and the value itself.

An important step is taken before proceeding into the calculation of the expected value. After calculating the vector A_p, a subset selection of the highest activation values is done, creating the vector A_{top_p} of n values with the highest activations. Given this, it is defined the new min_d and

max_d, as min_{O_d} and max_{O_d}, in order to re-normalize the data, similarly to equation 3.4, but only to the A_{top_p} values.

$$h''_{p,d} = \frac{h'_{p,d} - min_{O_d}}{max_{O_d} - min_{O_d}} \tag{3.7}$$

So, the expected value of dimension d can be calculated using all the normalized qualified values, which correspond to the h_d historical values that belong to $[min_{O_d}, max_{O_d}]$ and their activation values, calculated by the following equation:

$$E[O''_d|I''] = \frac{\sum_{p=1}^{n}(h''_{p,d} \cdot A_p)}{\sum_{p=1}^{n} A_p} \tag{3.8}$$

The variables in equation 3.8 correspond to:

- $E[O''_d|I'']$ is the expected value of the normalized output variable O'' under the normalized inputs I'';
- I'' are normalized inputs which were selected to calculate the expected value of dimension d normalized output;
- O''_d is the d dimension normalized output for which is being calculated the expected value;
- $h''_{p,d}$ is the normalized value of a p point in a d dimension, between $[min_{O_d}$ and $max_{O_d}]$;
- n is the number of qualified points.

For the calculation of the variance the second momentum calculation of the expected value is also needed. For that, the same equation is used, but with a slight difference, which is that the value appears squared, as following.

$$E[(O''_d)^2|I''] = \frac{\sum_{p=1}^{n}((h''_{p,d})^2 \cdot A_p)}{\sum_{p=1}^{n} A_p} \tag{3.9}$$

The calculation of the variance comes right after and is solely calculated using the expected values calculated before in 3.8 and 3.9.

$$V[O''_d|I''] = \frac{\sum_{p=1}^{n}((h''_{p,d})^2 \cdot A_p)}{\sum_{p=1}^{n} A_p} - \left(\frac{\sum_{p=1}^{n}(h''_{p,d} \cdot A_p)}{\sum_{p=1}^{n} A_p}\right)^2 \tag{3.10}$$

In order to simplify the analysis, equation 3.10 can also be written as 3.11.

$$V[O''_d|I''] = E[(O''_d)^2|I''] - E[O''_d|I'']^2 \tag{3.11}$$

After determining $E[O''_d|I'']$ and $V[O''_d|I'']$, obtained from 3.8, 3.9 and 3.11, αo_d and βo_d output values for each dimension d can now be calculated as described by the equations 3.12 and 3.13.

$$\alpha o_d = \frac{(1 - E[O''_d|I''])(E[O''_d|I''])^2}{V[O''_d|I'']} - E[O''_d|I''] \tag{3.12}$$

$$\beta o_d = \alpha_{d_o} \frac{1 - E[O''_d|I'']}{E[O''_d|I'']} \tag{3.13}$$

The outputs are then written the same way as inputs were, represented in 3.14, in order to further compare the difference between them.

$$O_d\{mino_d, maxo_d, \alpha o_d, \beta o_d\} \tag{3.14}$$

The following flow chart 3.3 summarizes the steps evidenced before.

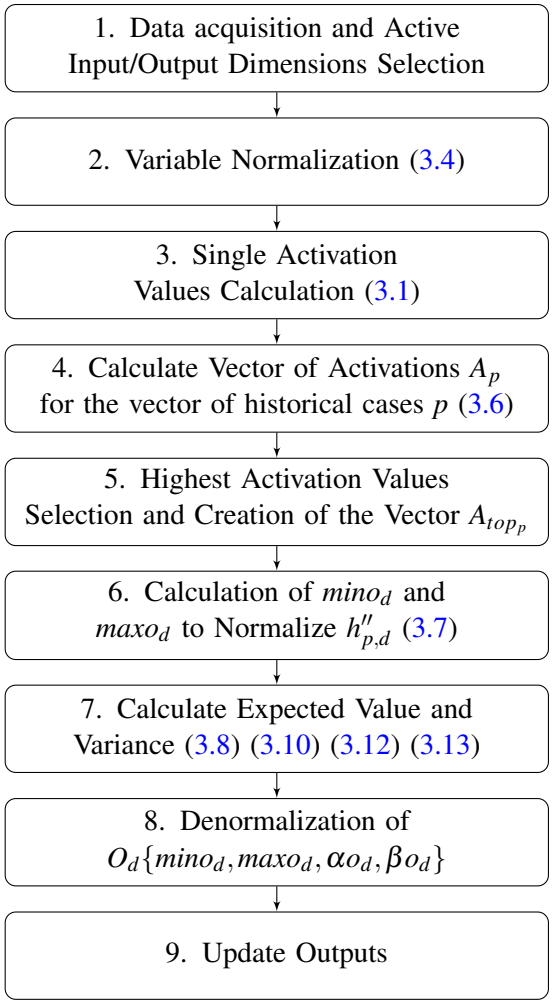

Figure 3.3: Proposed methodology approach

The method starts off with data acquisition (1.), where it gathers previously synchronized data and organizes it in a knowledge matrix, accounting for all the cases and dimensions. It also reads the information related to the dimensions, such as their parameters. After that, the user has the power to decide which dimensions will influence the output by turning on a binary code for the inputs. If a given dimension is turned on, then its parameters are relevant for the selection of qualified cases. At this point, it is filtered all reading errors and values outside the interval

parameters that was preset, by eliminating those cases from the knowledge matrix, allowing the method to work only under the desired values. It is important to emphasize that it is not necessary to ensure that the number of inputs to be equal to the number of outputs because the output variable is calculated using only activation values from qualified cases, since the activation values vector is independent from the number of variables, but dependent from the number of historical cases.

Step (2.) denotes the normalization done to the historical cases under equation 3.4. This is a very important step since a bad normalization can result in a poor application of method or it can even cause some fatal errors to emerge, such as the values being outside the allowed interval of $[0;1]$.

Next, it is (3.) calculated the single activation values for each historical case p of each dimension d, resorting to equation 3.1, whose inputs are $I'_d(h'_{p,d};\alpha i_d;\beta i_d)$, which account for the input I'_d historical value p of dimension d and the corresponding αi_d and βi_d. As it was previously explained graphically by figure 3.1, different values of α and β parameters have different effects on the calculation of the activation values. For example, setting α and β equal to one would assign the same activation to each value of a given inteval. By setting α and β to higher values, for example, equal to two, on both parameters, the values that appear on the central part of the interval will have higher activation values, contrarily to the ones near the minimum and maximum limits.

Once data is saved in the knowledge matrix, already filtered, normalized and transformed into single activation values, the next step takes on the calculation of A_p, a joint activation vector (4.). The method starts by calculating A_p through 3.6 for the corresponding case and stores the value in the correct position.

The variation presented in this method comes in the form of a filter (5.), which does not allow the user to apply the method without enough historical data to correctly identify the Beta PDF parameters. Considering this, after a first evaluation of the historical sets that qualify for the application of the method, that is, after calculating A_p, the joint activation vector is sorted from highest activation to lowest activation value. At this point, the selections is done through two different means: through user choice, as the user has the power to select the number of cases, or through minimum level of activation values. The last one, the minimum level of activation values, contemplates a percentile of the total cases, eliminating the ones below a certain level, as the remaining values constitute the final vector A_{top_p}, to be used next in the calculation of the expected value and variance. The minimum level is defined at the start of the program and, in this dissertation, it was decided to be 10% of the maximum activation of A_p. The user is allowed to change the value as pleased.

This way, the user has the choice to use every single value of the qualified cases, or only use the ones with the higher activation values, which are closer to the kernel of the entire data set of the selected variables. This allows the cut of some dispersed values which could add noise to the final results, turning the method more accurate. It is important to emphasize that the interaction with other external agents is done at this point through the sharing of the activation vector A_{top_p}.

In short, the interaction with other agents, which in this case will be electrical nodes, is done, as stated before, through the sharing of of the activation vector A_{top_p}. This was emulated in a simple

Excel file, which served as a cloud service, in which every node sent their activation vectors for each qualified case. The updated information about the weights of every case could be later pulled by other agents, without knowing the actual information, in order to calculate a column vector of the activations for each case.

After calculating the subset of the highest activation values, a (6.) determination of the new maximum and minimum values is needed, as well as a re-normalization of the historical values. Since the method previews the use of an interval between $[0,1]$, the fact that the selection may have shrunk the interval, a new normalization is needed, using equation 3.7. In order to do that, the vector A_{top_p} is first read and then it is identified the positions that match the activation values to the actual values. It is then performed an evaluation to determine the new $mino_d$ and $maxo_d$, enabling the application of the equation referenced before.

Once the qualified values are selected, equations 3.8, 3.9 and 3.10 are applied (7.) to all the active output dimensions.

A problem was encountered as αo_d and βo_d were lower than 1. Such values turn the Beta PDF into exponential ones, matching higher probabilities to values near 0 when αo_d is very low, and matching higher probabilities to values near 1 when βo_d is very low. Since those results are incorrect and do not reveal the true nature of the data, a small verification was made just before calculating both αo_d and βo_d. The verification to ensure that αo_d and βo_d are never less than 1, equations 3.15 and 3.16 were used. There was also a need to verify the variance value through equation 3.17, in which *NMP* is a very small number, in order to not only limit smaller values of αo_d and βo_d, but also the greater values of them too, as that situation is also problematic, since it assigns very high probabilities (near 1) to a limited interval and very low (near 0) to all the remaining interval.

$$\text{If: } V[O''_d|I''] > \frac{E[O''_d|I'']^2(1-E[O''_d|I''])}{1+E[O''_d|I'']} \quad \text{Then: } V[O''_d|I''] = \frac{E[O''_d|I'']^2(1-E[O''_d|I''])}{1+E[O''_d|I'']} \tag{3.15}$$

$$\text{If: } V[O''_d|I''] > \frac{E[O''_d|I''](1-E[O''_d|I''])^2}{2-E[O''_d|I'']} \quad \text{Then: } V[O''_d|I''] = \frac{E[O''_d|I''](1-E[O''_d|I''])^2}{2-E[O''_d|I'']} \tag{3.16}$$

$$\text{If: } V[O''_d|I''] < NMP \quad \text{Then: } V[O''_d|I''] = NMP \tag{3.17}$$

Finally, output $mino_d, maxo_d, \alpha o_d$ and βo_d are stored in the correct output active dimensions (9.), ending the application of the method. Before being loaded into the spreadsheet, a (8.) denormalization is needed applying the inverse method. The output values are then ready to be used in probabilistic evaluations.

There are two important aspects of the methodology described above that need to be emphasized beforehand. The first one is that this is not an iterative method, as the outputs can not turn into inputs, at least α and β values, since it will simply alter the single activation values for the

different historic cases of a given input variable. This will shrink the curve as it is mathematically implying that the user has some knowledge about the behaviour of the tested variable.

The second one is a consequence of the first: as the user shall not know the behaviour of a given variable, initially at most, said variable shall be initialized having α and β equal to 1. This allows the software to assume that any given value is as likely as any other value present in the historical data, which means A_p will only be influenced by the concentration of points around a certain value.

Essentially, the communication between neighbour buses can done as soon as step number 3 of the flow chart is completed, that is, the input variables qualified sets change their single activation values, which in turn will change A_p for every qualified set.

This communication becomes safer for other neighbour buses to use, as it does not take unnecessary risks by handling important and explicit electrical information, but instead it simply passes the joint activation vector, encoded by default since it denotes the multiplication of several single activation values of each variable for a specific case from the entire historical set.

3.2 Explanatory Example

This section has the purpose of explaining some results derived from the methodology presented previously. The tests presented in this section contemplate one input dimension and one output dimension, in order to facilitate the comprehension of the results shown later in chapter 5. For that set, it is evaluated the Reliability Index, as well as it is done a brief comparison between the real test value and quantile Q_{50} value.

In order to do that, the data was firstly divided between training set, also known as historical data, and testing set, which corresponds to new input values. The only difference between real application and test is that the output real value is given to be evaluated further in this dissertation.

Firstly, the input distribution is set by default as α and $\beta = 1$, $min = 0$ and $max = 1$. Then, the user has a choice to set the range of the several inputs by changing the interval value. This interval value is composed by a random number generator, which shifts the interval around the input value for the given case. From there on, that value is chosen as reference and the random number will create an interval, not centered in the real value, but containing it. The example shown in this section is using an interval of an amplitude of 0.05.

After setting the input alpha, beta, maximum and minimum values, the data is then set to be applied to *DenDEx*. The method first loads the qualified cases of the knowledge matrix, that is, the ones that comply with the interval set for input variables, creating the vector that contains the available data. At this point it is done a standardization between a very small number and one minus the very small number. This is done to avoid near zero values for the distribution, whose problems were already identified previously.

Then, it is calculated the activation function by 3.6 for each entry, that is, for each qualified case. The comprehension of the meaning of this value is of utter importance. Finally, the method

ends by printing the output parameters. This method is applied to the entire test set, which contains 768 different values, corresponding to different situations from different chronological instants.

A reliability diagram can be a great indicator of the "reliability" of the method, whose purpose is to assess the frequency of a given real result to fall in a specific interval, hereinafter called quantile. The quantiles are supposed to divide a given distribution between a specific number of equal areas, or probabilities. In this dissertation, the normalized interval between zero and one will be divided into quantiles, intervals of 0.1, totaling an amount of ten intervals. Therefore, the ideal frequency, or target frequency, of each quantile is 10%. As it is observable in figures 3.4 and 3.5, the real value tends to fluctuate between quantiles Q_{10} and Q_{90}. Also, figure 3.6 compares the target frequency to the observed frequency.

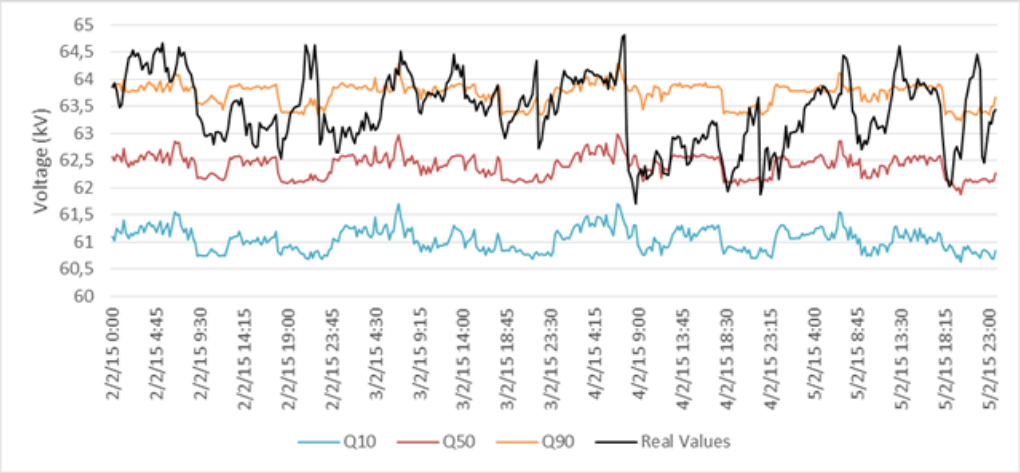

Figure 3.4: Winter set example results

Figure 3.5: Summer set example results

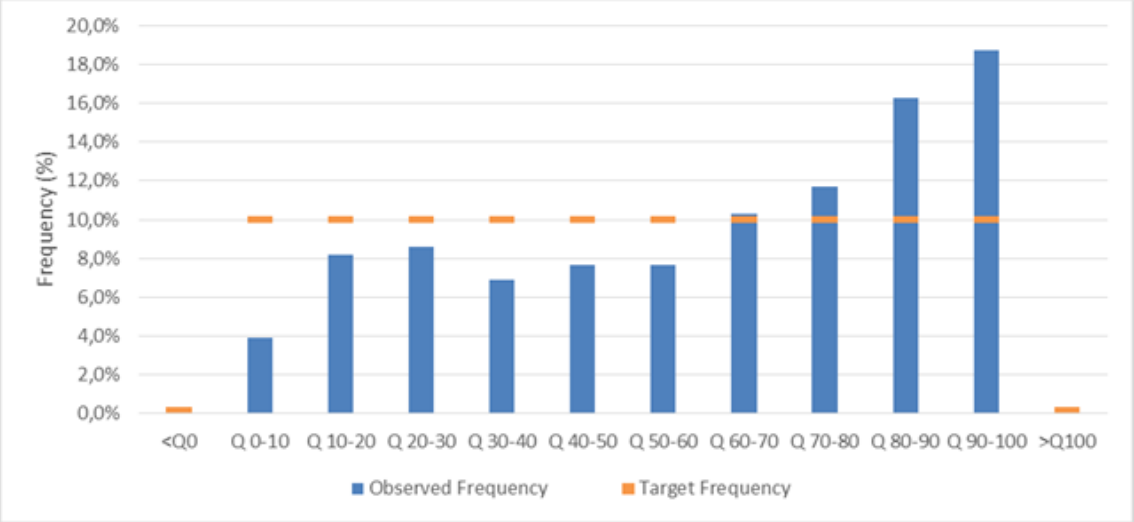

Figure 3.6: Quantile distributions of example results

In order not only to evaluate the results visually, it was used several indexes to correctly identify the accuracy of the method.

Firstly, a reliability index *RI* was created, which calculates, resorting to equation 3.18, the average normalized difference between calculated frequencies and the target frequencies. It is important to emphasize that in order to calculate *RI* a test set is needed, since each case corresponds to one result. To evaluate the previous index it is needed a great number of test cases to be able to verify the distribution.

$$RI(\%) = \left(1 - \sum_{i=1}^{NQ} |f_{obs_i} - f_{target_i}|\right) * 100\% \tag{3.18}$$

in which:

- f_{obs_i} is the observed frequencyof quantile i;
- f_{target_i} is the target frequency of quantile i;
- NQ is the total number of quantiles.

According to equation 3.18, higher values of *RI* means closer values to the target frequencies. The example shown previously has an *RI* of 65.9%.

Secondly, it was used the *Mean Absolute Percentage Error - MAPE* - using equation 3.19. MAPE is a very important index in forecast algorithms to assess the accuracy by a means of a percentage, comparing a target value to an output value. In this case, the target value is the real value of a test case and the output value is the Q_{50} quantile value.

$$MAPE(\%) = \frac{100}{n} \cdot \sum_{i=1}^{n} \left| \frac{T_i - O_i}{T_i} \right| \tag{3.19}$$

in which:

- T_i is the target value;
- O_i is the output value, which corresponds to Q_{50} quantile value;
- n is the number of cases.

Contrarily to *RI*, lower values of *MAPE* index means the output values are closer to the target values. The example shown previously has a *MAPE* of 1.38%. Another index of a possible evaluation is the average difference between quantiles Q_{10} and Q_{90}, which in this case was determined to be 6,296 kV.

The next chapter presents an overview of the information used and how it was handled in order to prepare it to be applied to this method.

3.3 Connectivity, Reaction and Synchronism

In this section will be presented one idea that appeared evident during the writing of this dissertation, but that could not be tested. One very important aspect of this dissertation that was lacking was the capability of communication. This whole idea is the foundation of something bigger, that could be applied basically to everything that handles correlated information. The fact that there is a need for communication raises some concerns about security, which takes to the second aspect that needs work: information transaction.

Instead of running all variables at the same time, imposing the data transaction, a better idea was found, which contemplates the transfer of encoded information via activation values. This would allow the users to use valid information without the need of directly handling sensitive information.

This way, the raw information is protected from piracy or nefarious effects and the program is still allowed to run, since all information is encoded in the activation values for the given input dimensions. From this new proposed methodology, it can only result in benefits, from the privacy and computational time point of view.

Given the stated above, all of the methodology's objectives are the same, but the process contemplates slight differences. The method above is thought and is operated having information regarding the whole knowledge matrix. That enabled the users to directly use information from other buses, which compromised the security of the data. This way, altering the method under the differences between figures 3.3 and 3.7, would provide not only more security, but faster interactions between information of different dimensions.

The process of calculation of vector A_p, which is a column vector corresponding to the joint activation values of a given case for all the input dimensions, is slightly different.

The differences start by the selection of input dimensions. For example, assuming only two dimensions, one input and one output. As soon as the entities that control the information regarding those dimensions, receive updated values, they can calculate the single activation value of the different measures, through equation 3.1.

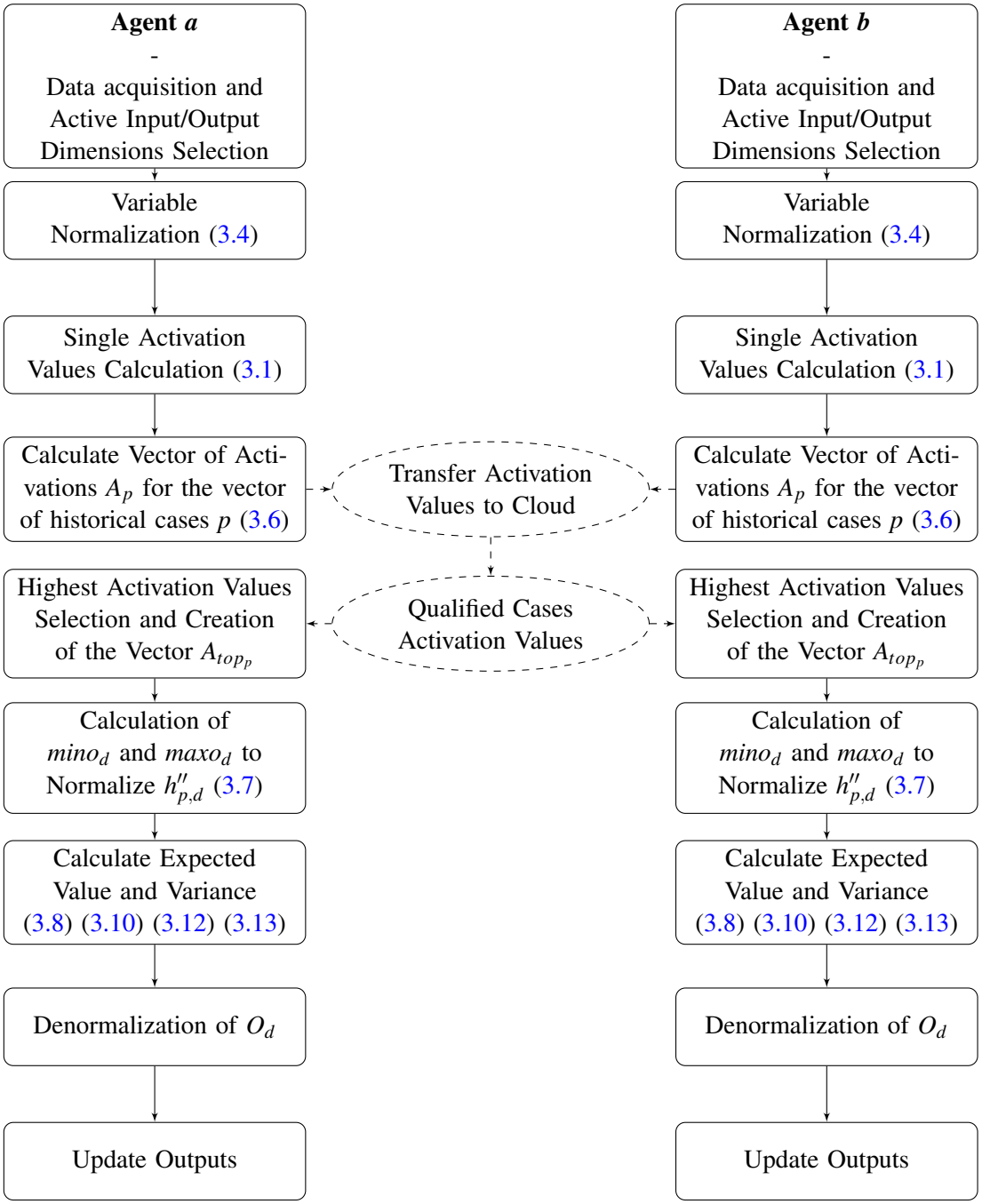

Figure 3.7: Alterations to the proposed methodology

After that, those activation values are sent to a cloud environment, through a public key encoded with the hour and date of the historical case, the hour and date of the forecast instant and

a private code of the simulation. This blockchain coding allows a public access to the activations for a specific predictive instant coded in the blockchain. The agent simply queries the cloud for all registers available for the target predictive instant, getting registers from multiple simulations activating multiple historical cases, with identifiable date-time codes. Some information received through the request is useful and some other is not, so it will be on the agent side that the decision process to select the information to be used is going to be made. Based on the activation correlation between the registers of the agents' simulations and the registers of other simulations received from the cloud, the agent dynamically identifies the external simulations and correspondent activations that are useful to enrich the knowledge matrix of their on simulation. The matching of registers of internal with external activations is done in the user side, as the cloud only acts as a storing device for the blockchain registers, without any kind of relation or iteration between registers.

Given the represented in figure 3.7, a briefly explanation is handled. Assuming agents a and b are from two different entities and are holding different kinds of information and two cases are verified: the first one, information from both dimensions is synchronized and is valuable; the second one, information is not synchronized or is invaluable.

For the first case, the proposed methodology in 3.3 acts perfectly, since the use of all information is based on the premise that it is synchronized, hence the ability to push information is the same as having all the information from agent b known on the a side and vice versa. It is calculated the joint activation value vector A by equation 3.6 for the two dimensions, it is then selected the highest activation cases, which is then applied to the expected value and variance equations.

However, assuming that this premise is not fulfilled, the information is not synchronized and the selection of viable cases and input dimension can not be done as one pleases. If there is no matching information from agent b, then the output variable can only be forecasted through the use of agent a self information, which is not sufficiently good to have a decent forecast. Now, assuming that there is not only two agents, but one hundred, the probability of asynchronous values rises.

The idea is then to confer the power of decision to the different agents on which input variables and qualified cases they should use. After agent a calculates its input activation values, through equation 3.20, it passes that encoded information to the cloud environment.

$$A_{p,entity} = \prod_{d=1}^{m} I'_d(h'_{p,d}; \alpha i_d, \beta i_d) \tag{3.20}$$

The difference between equations 3.6 and 3.20 is that equation 3.20 solely calculates one agent's activation joint values. This allows the information's activation values to be ran through the cloud and then multiplied after the selection, which turns equation 3.6 into the product of multiple entities' activation values.

Here, the cloud determines which inputs from other agents are available for that instant in particular and after selecting the qualified cases, it sends back those values to the different agents. The agents can then select the activation values as they please, that is, they can assess themselves

if the information is useful or not. One suggestion is to evaluate the correlation between activation values. Higher correlation between variables mean that they influence each other in a greater factor than others with lower correlation variables.

After that, the agent a gathers the cloud information, which contains the activation values of the cases of several agents, it is ran the method as proposed, by selecting the highest activation values, calculating the expected value and variance and, ultimately, updating their outputs.

Chapter 4

Case Description

This chapter will address the data available with an extensive description of its management and usage.

4.1 Case Considerations

Management and control of electrical grids, both transmission and distribution, are operated individually by separate companies. Transmission grids are operated by the *Transmission System Operator*, also known as *TSO*, which in Portugal is concessioned to *Redes Energéticas Nacionais - REN*. However, there is also the operation of distribution networks, smaller in voltage and power levels, but more numerous, which are operated by the *Distribution System Operator*, or *DSO*, in which the vast majority of the distribution grids are concessioned to *Energias de Portugal - EDP*.

The data used in this dissertation is from a distribution network from *EDP*. In order to respect confidentiality agreement, the names of the buses will be concealed and changed to numbers.

As it is known, *EDP* has two major centers of operation of distribution systems: one located in Porto, which controls the North region of Portugal and another one located in Lisbon, which controls the Center and South regions. Inside those operation centers, it is also divided by voltage level, having different teams for high voltage and medium voltage grids.

Given this, the data used is the same as used previously by those control centers, being handled by a *SCADA* software. Incorporated in that software there are functions which retrieve and display the information sent remotely from the IED's located in the substations. That information involves both electrical data such as voltage profiles, currents and apparent power injection, as well as topology status, which contemplates information about disconnectors, whether they are open or closed, state of operation of capacitor banks, transformer taps, among other relevant information such as temporal data.

This allows the teams to have a deterministic overview of the current status of their grids and how they are being operated. This data is also useful in terms of planning, since there is already some knowledge about which branches are maxed out, and how the production/loads are distributed.

The grid at study is particularly interesting due to the presence of renewable sources, such as hydro and wind power, as well as connections to transmission grids. This allows to have very different profiles from case to case, which is a good starting point, at least for the country energy mix repertoire.

Portugal is a rich country in hydro resources, consolidating hydro power as one of the most important sources of renewable energy. In the last decades, there has been a growth, due to the magnificent incentives given by the country's government, of wind power sources, which propelled the construction of several wind farms. The following figures 4.1 and 4.2 are a keen representation of the presence of wind energy in Portugal's energetic mix and the growth over the years.

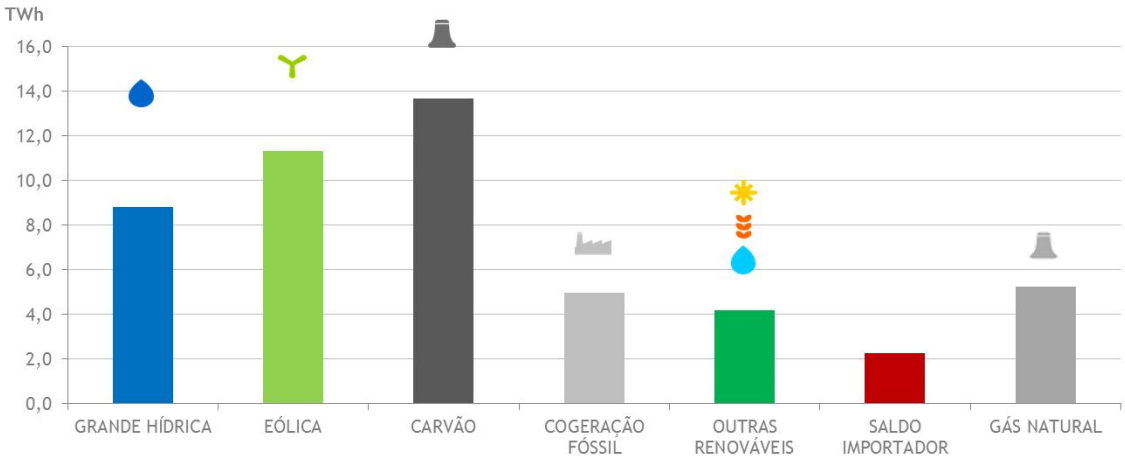

Figure 4.1: Portugal Energy mix, from January 2015 to December 2015. Adapted from: [46]

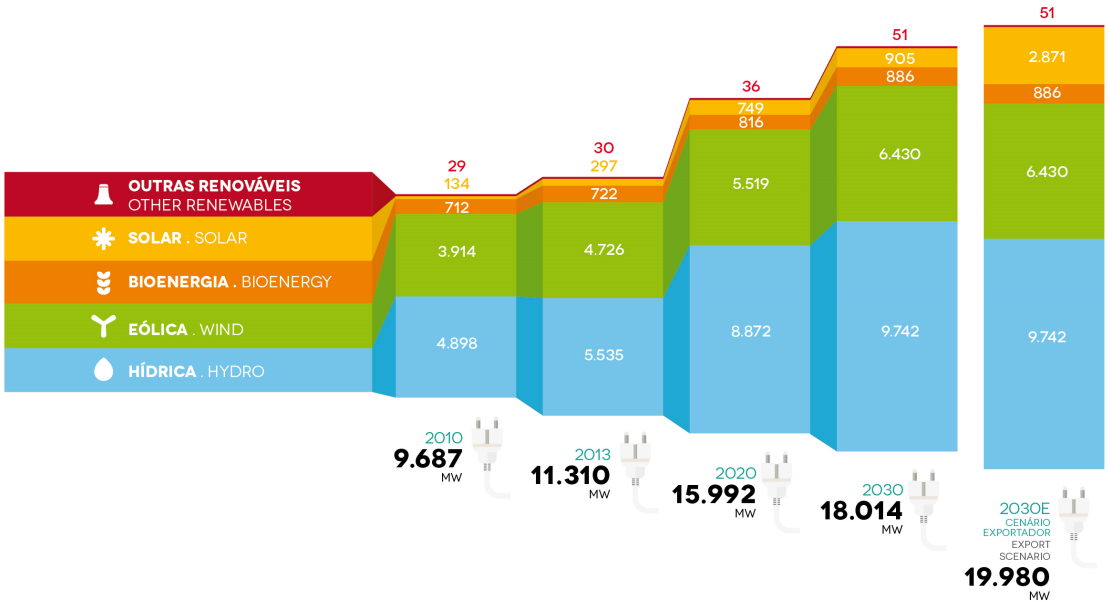

Figure 4.2: Portugal Renewable Energy mix growth. Adapted from: [47]

The strong growth of wind energy production allowed to shift storage hydro power from the load profile base to peak, that is, since the majority of the wind energy production is during the night time, in valley hours, that energy is used to fill the dams' reservoirs to respond later in peak hours.

Besides that, the rise of wind energy presented itself as a challenge to system operators, since there is high variability in power output from hour to hour, forcing the operation to account for rapid variations of power injection in several buses.

This grids' knowledge matrix is then rich in cases, from wind farms to hydro power plants and loads, setting itself as a great test for the tool developed. The next section describes thoroughly the data involved in this dissertation.

4.2 Data Analysis

The methodology discussed previously in this dissertation revolved around a real case scenario, being the data from a 60 kV grid. The grid information will be, from this point forward, protected by using code names for the various buses, as stated before.

The following figure 4.3 is a representation of the grid's single line diagram.

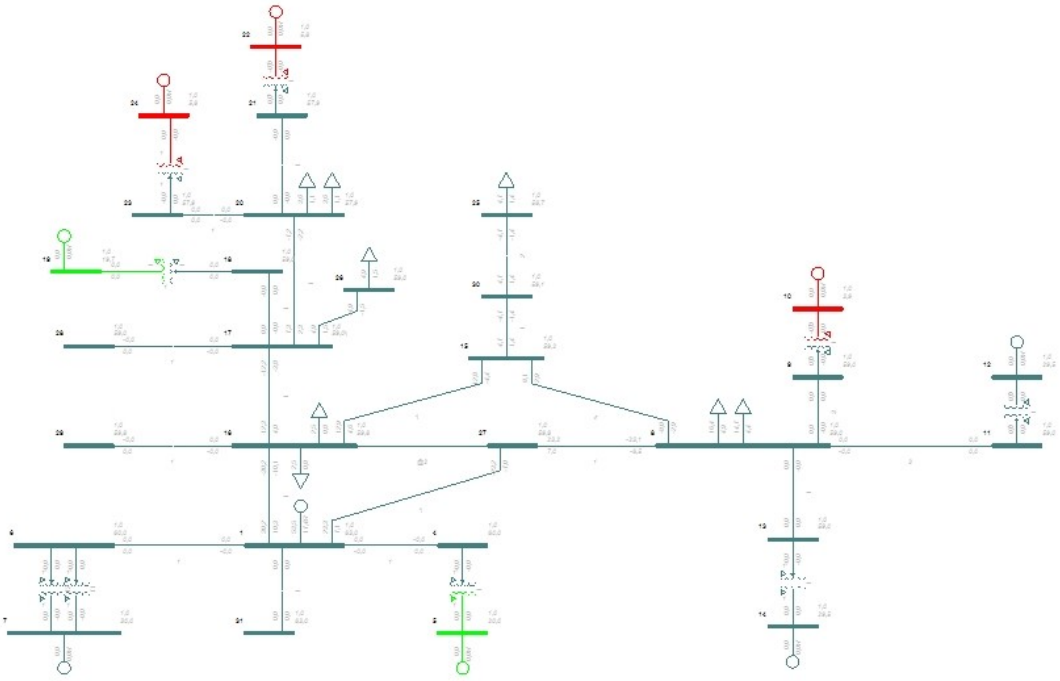

Figure 4.3: Real case scenario single line diagram

The grid at study is operated in open ring topology and is composed by a total of 31 buses, 9 generators and 8 loads, corresponding to the calendar year of 2015, each measure being taken fifteen minutes apart, making a total of 34272 cases. As the information was very fragmented

and unsynchronized, which was very difficult to handle, a small code was developed to overcome blank cells and repeated values, which was different for each file of the four electric units related to each bus, in order to organize with no duplicates to easily synchronize all measures from near buses.

The available data prior to the organization was the following:

- **Temporal Data:** Day, day of the week, month, year, hour and minute of data acquisition;
- **Electrical Data:** Voltages, currents, active and reactive power;

  Bus 1: Voltage on LV side;

  Bus 2: Voltage on HV side, active/reactive power injection;

  Bus 4: Voltage on HV side, active/reactive loads;

  Bus 5: Voltage on HV side, active/reactive generation;

  Bus 6: Voltage on HV side and active/reactive loads;

  Bus 61: Voltage on HV side and active/reactive generation;

  Bus 62: Voltage on HV side and active/reactive generation;

  Bus 7: Voltage on HV side and active/reactive load;

  Bus 8: Voltage on HV side and active/reactive load;

  Bus 10: Voltage on HV side, active/reactive generation and active/reactive load;

  Bus 11: Voltage on HV side and active/reactive load;

  Lines L41, L54, L56, L105, L43, L49, L91: Currents in one or both directions.

Some of the grid components did not have available data such as some wind farms located in Bus 1 and Bus 2.

Bus 1 is one of the most important substations of the network, since it corresponds to the injection point from the transmission grid. Normally, the voltage in this bus is very stable and above nominal voltage, in order to take on voltage drops throughout the grid. However, this is on of the buses with several data errors, not being included in the analysis made in the next chapter. Due to the high reactances of the branches, many times greater than their resistances, injected reactive power has a great effect on bus voltages.

As for an example, the bus selected to do the vast majority of the tests was bus 6, located in the periphery of the grid, being only connected to two hydro power plants and bus 5. The following figure 4.4 presents bus 6 voltage, active power and reactive power on the course of an year.

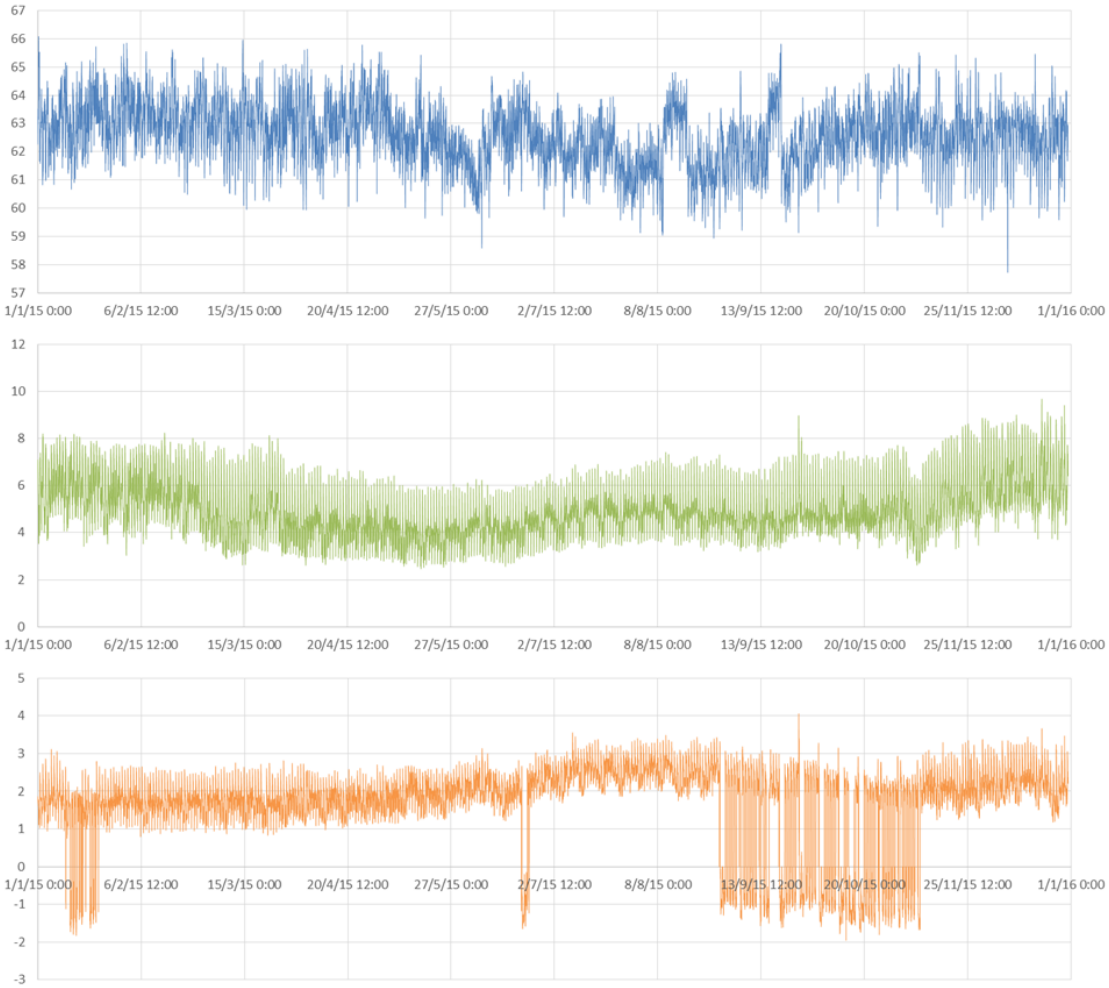

Figure 4.4: Bus 6 Voltage in kV (Blue), Active Power Injection in MW (Green) and Reactive Power Injection in MVAr (Orange)

After organizing all information by bus and by electric unit, it was performed a synchronization of the data in order to have in singular *Excel* files all the data related to the nearest buses.

The next step was to correctly identify the possible topologies in which the grid was operating. By studying the available data, it was possible to state that the grid is normally explored by disconnecting itself from the upper grid (bus 10) and the lower grid (bus 11). Two other scenarios were found: the first involved the disconnection of line 41, which links bus 1 and bus 4, turning the power supply to lines 43, 49 and 91, which is, physically speaking, a triple junction between buses 1, 2 and 4 and the second both lines 41 and the triple point are out of service, which interrupted the power supply from the transmission lines. The last one only happened 16 times during the year, accounting for a total of four hours.

Finally, to completely characterize the data, it was retrieved two weeks, presented next, with particular parameters identified below. The objective of this selection is to validate results from the methodology previously presented in this dissertation. This selection, which is represented in

figure 4.5 can be described as:

- **Winter example set**

  Number of records: 384;

  Days of the month: 2 to 5;

  Days of the week: Monday to Thursday;

  Month: February.

- **Summer example set**

  Number of records: 384;

  Days of the month: 8 to 11;

  Days of the week: Monday to Thursday;

  Month: June.

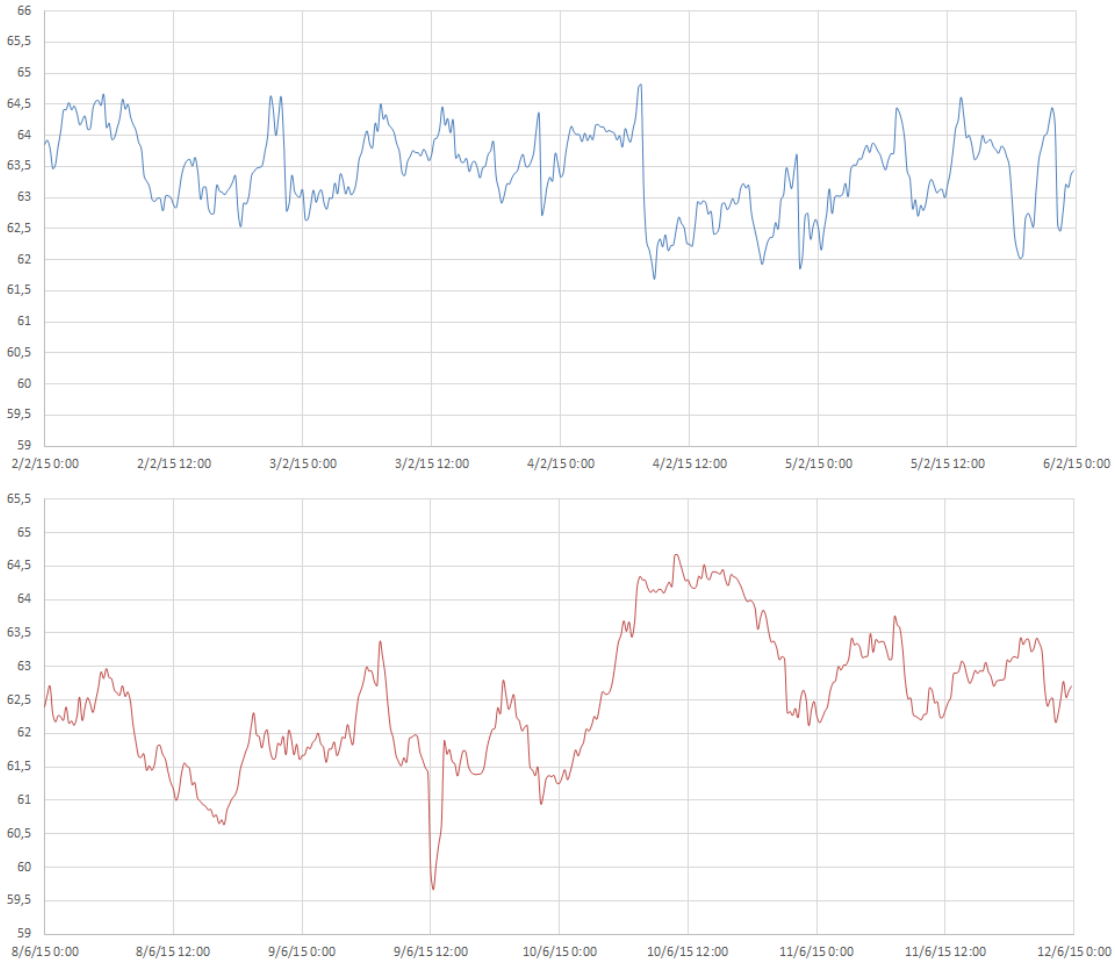

Figure 4.5: Bus 6 Winter Set Voltage (Blue) and Summer Set Voltage (Red) in kV

Following inter-bus organization of the data, it was decided to evaluate the variables, either temporal or electric, under Beta PDFs. The aim is to use Beta distributions as standard models of said variables when there is not enough historical data available to perform the PLF method.

Given the stated previously, the first step, after synchronizing the information, was to normalize it between 0 and 1. In order to do that, 3.4 was used.

After normalizing all the variables, and as a means to apply the proposed method, it was created three blocks (standard block, input block and output block) with four parameters each (α, β, *min* and *max*). The standard block parameters were filled with the results of a study made to the behaviour of the variables on the course of a year. The study consisted in the evaluation of the number of times a given value occurred within a given range.

Firstly, it was defined a range between 0 and 1 and it was then tested the division in different intervals. The best results were achieved by creating intervals with an amplitude of 0.025. It was then calculated the number of values within each range, which were then divided by the total occurrences, as a form of probability of occurrence, so to speak.

Secondly, the way to correctly identify the Beta PDF was to use the *Excel Solver* routine to minimize the quadratic error between the probability of occurrence, calculated before, and the subtraction between the CDF in the beginning of the interval and the end. The restrictions were to α and β, which were restricted to $]0,-\infty[$. The initial conditions were flat-starts with $\alpha = \beta = 1$ and *min* and *max* equal to 0 and 1, respectively.

All standard Beta PDF were determined by this method. The only exceptions were to the variables which *Solver* could not find a suitable solution, satisfying all restrictions. In those cases, both standard and inputs were initialized as $\alpha = \beta = 1$ and *min* and *max* equal to 0 and 1, respectively.

Chapter 5

Simulation and Results

The base case is composed by 6 buses, from the total 31 buses. This is due to the lack of quality from the data, which often contains errors which makes some of the buses impossible to use. The buses used were B4, B6, B7, B12, B61 and B62. This selection of buses enabled the evaluation of certain aspects of HV grids, since B6, a bus with active and reactive loads, is directly connected to B61 and B62, which hold two generators from hydro power. Also, although a little far away from these, B12 holds a wind farm and B4 and B7 more active and reactive loads, which are not connected to any of the generation considered before.

In order to limit the number of figures presented in this and the following sections, it will be evaluated the most relevant cases, mostly related to buses B6, B61, B7 and B12. The reasons behind this selection are the following: B6 presents loads near hydro generation, B7 presents loads far away from any generation, B61 represents the behaviour of hydro generation and B12 the behaviour of wind farms.

It is important to emphasize that the results shown in terms of Beta PDF Distributions were randomly selected from an extensive test case, which turns them into just illustrative examples. The conclusions drawn refer to the entire test set.

5.1 Base Case

This is considered the base case, since it used only active power from the six buses to estimate the voltage outputs of those same buses. The chosen interval of inputs had an amplitude of 0.5 around the input values. This interval can be justified by the number of input variables involved, as a too smaller number would not return any qualified cases.

In this section will be evaluated the values of output voltages, output behaviour of the inputs, output reliability, output *MAPE* and output average difference between quantiles Q_{10} and Q_{90}. After choosing active power of different buses as inputs, for each one of the 768 test cases, the results were as follows.

5.1.1 Bus 6 - Load Bus Connected to Generation

The analysis starts by investigating one of the central buses from the selected ones, as B6 contemplates active and reactive loads, as well as being close to a power source. Figures 5.1 and 5.2 evidenced that the output distribution is closer the real values of the voltage in the winter, rather than in the summer.

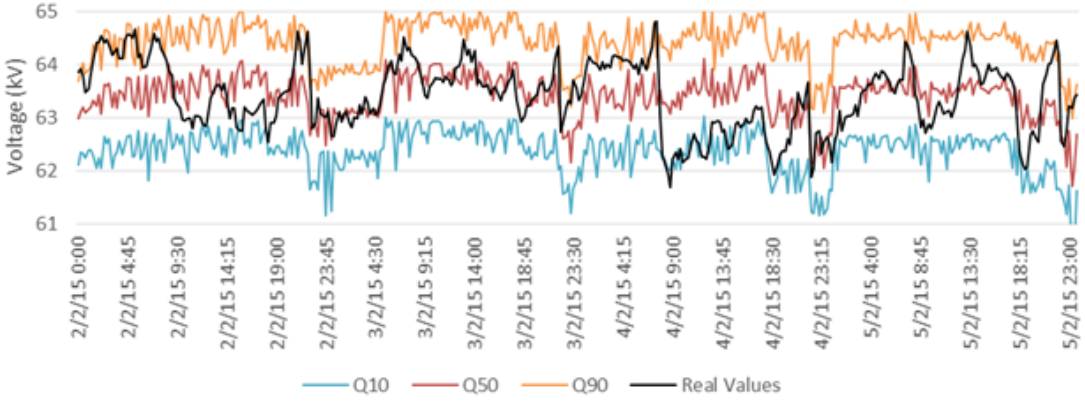

Figure 5.1: Winter Set B6 Voltage

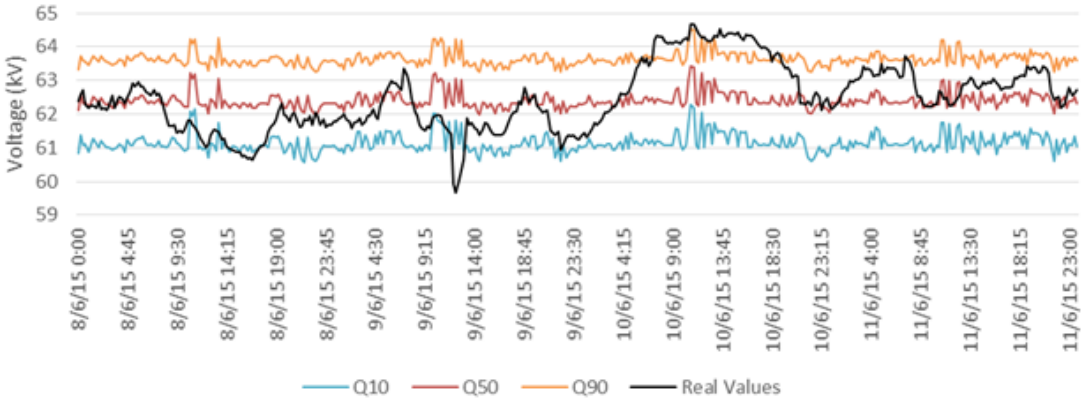

Figure 5.2: Summer Set B6 Voltage

By analyzing carefully figures 5.1 and 5.2, one can take some conclusions:

- Winter set quantile values are led significantly better by the real values than the summer set;
- Winter set presents some voltage drops between the 2^{nd} and the 3^{rd}, the 3^{rd} and the 4^{th} and also the 4^{th} and the 5^{th} of February;
- Summer set quantiles present some regularity, contrarily to B6 voltage real values.

Voltage drops in figure 5.1 can be explained by sudden drops of active power production in B61, evidenced in figure 5.3. Opposing to that, voltage rises displayed in figure 5.2 are directly linked to the peaks of active power production presented in figure 5.4.

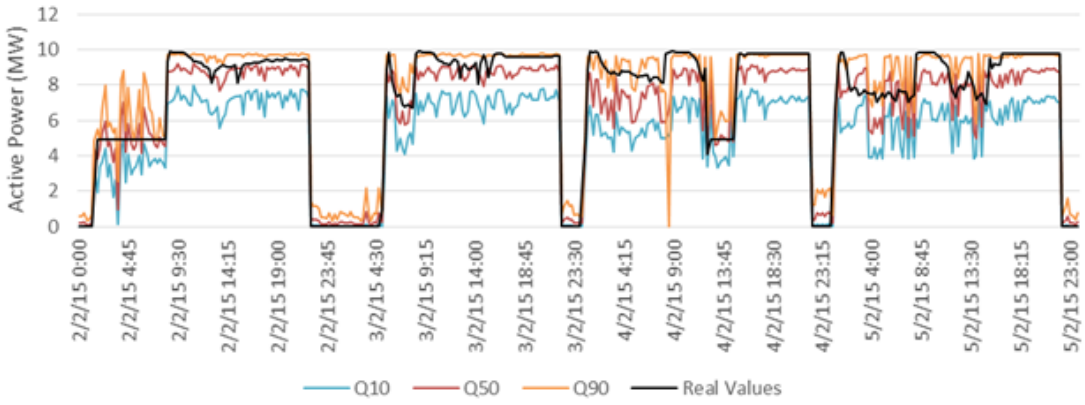

Figure 5.3: Winter Set B61 Active Power

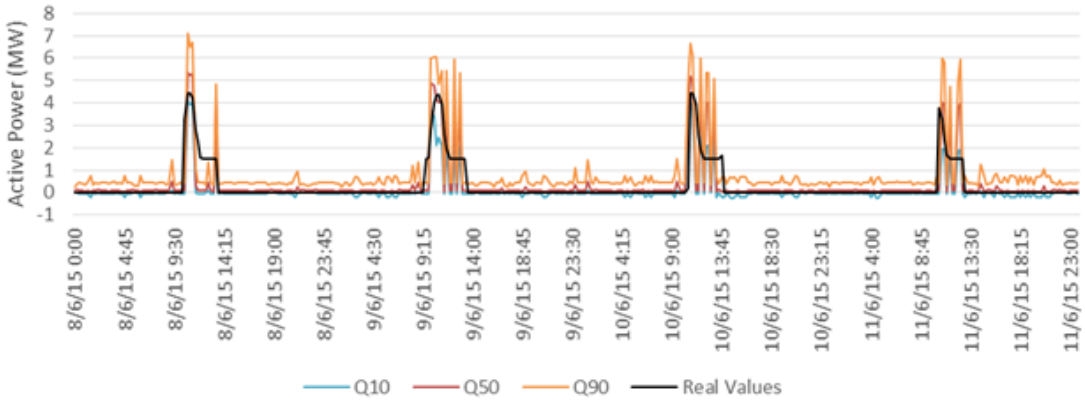

Figure 5.4: Summer Set B61 Active Power

This results translate into a distribution by the ten quantiles as presented in table 5.1.

Table 5.1: Quantile Distribution of B6 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
6,3%	8,6%	11,7%	10,2%	11,7%	11,9%	11,9%	9,4%	8,5%	9,5%

As it as evidenced before, table 5.1 can be represented graphically for a better understanding. Figure 5.5 shows the absence of values outside the bounded interval. Quantiles Q_{20-30}, Q_{40-50} and Q_{60-70} are almost 2% above the 10% target frequency and quantiles Q_{10-20} and Q_{80-90}, almost 1% below the target frequency.

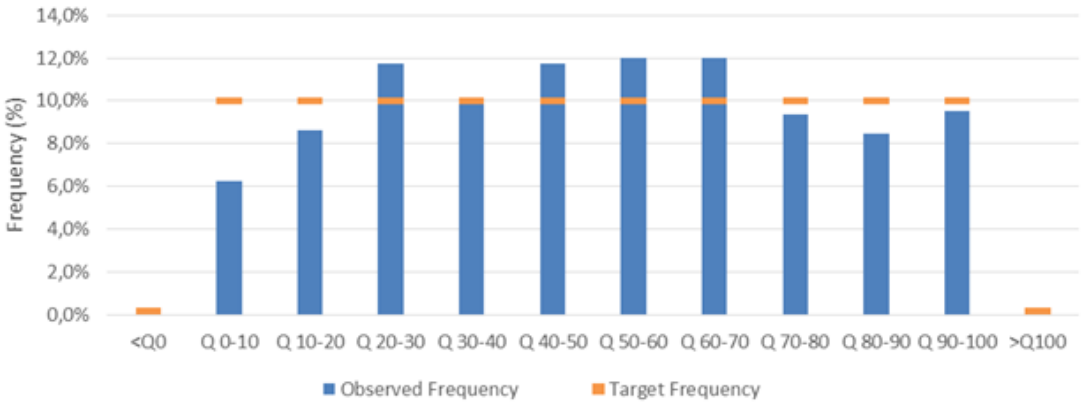

Figure 5.5: Quantile Distribution of B6 Voltage

After calculating the distribution by quantile, it was calculated the *RI* and *MAPE*, which returned a value of 84,6% and 1,01%, respectively. Another index was evaluated solely for B6 which contemplates the average difference between quantiles Q_{10} and Q_{90}. For the base case the difference was of 2,27 kV. This difference, having 60 kV as per unit (p.u.) base voltage, represents an interval of 0.038 p.u., which deems itself acceptable.

5.1.2 Bus 7 - Load Bus Far from Generation

Similarly to B6, B7 is also characterized by having active and reactive loads, but further away from any power source. Figures 5.6 and 5.7 evidenced that the output distribution is not as influenced as the previous bus, but still corresponds to slight variations of real values.

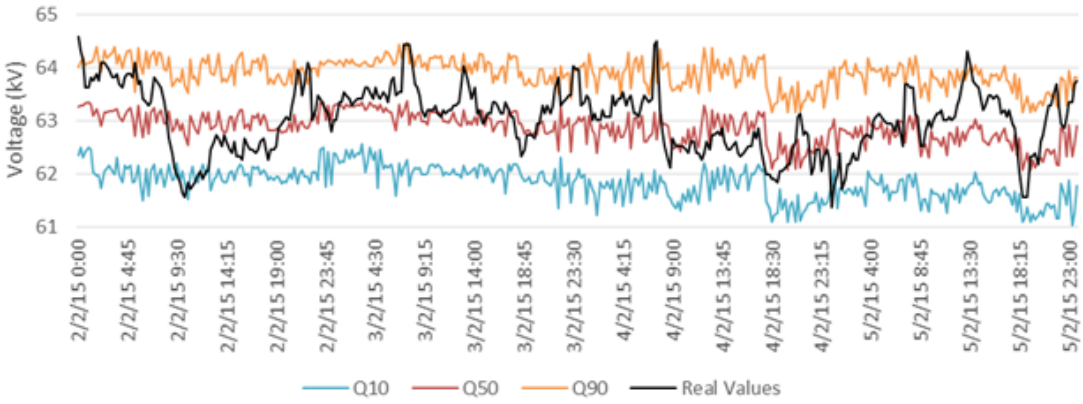

Figure 5.6: Winter Set B7 Voltage

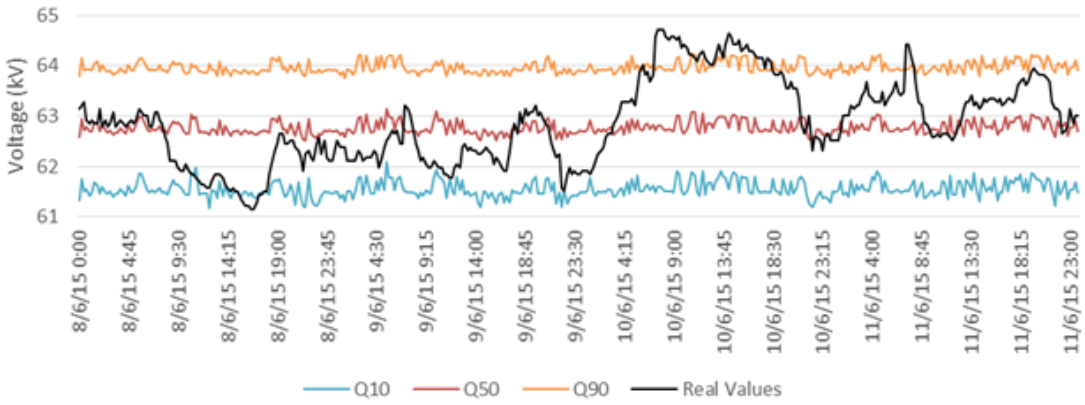

Figure 5.7: Summer Set B7 Voltage

By analyzing carefully figures 5.6 and 5.7, one can take some conclusions:

- Winter set quantile values are still led better by the real values than the summer set;
- Both Winter and Summer set quantiles present some regularity, contrarily to B7 voltage real values.

The regularity presented, with predominance in the summer set, but also observable in the winter set, is due to the lack of proximity to any power source. As voltages are susceptible to power variation and this bus is far away from any power source, the ripple caused by those power variations tend to fade. Another aspect that helps the fading of those variations is the fact that the input interval is such that any conjugation of qualified cases picks very regular voltage values.

Figures 5.8 and 5.9 shows the quantile distribution of the input variables, to better understand if the method correctly identifies the tendencies of the said input variables. As it can be observed in 5.8, there is a certain pattern, where one can see the rise in active power consumption around 19:00 everyday, potentially denoting residential loads.

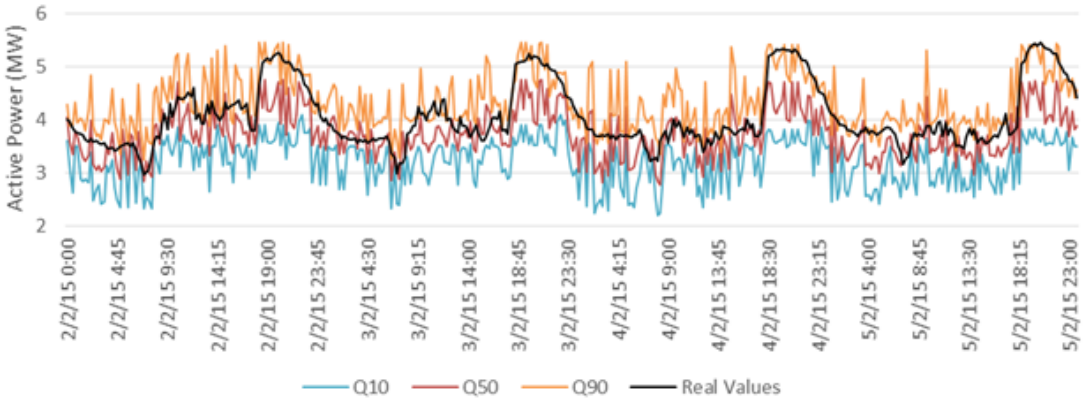

Figure 5.8: Winter Set B7 Active Power

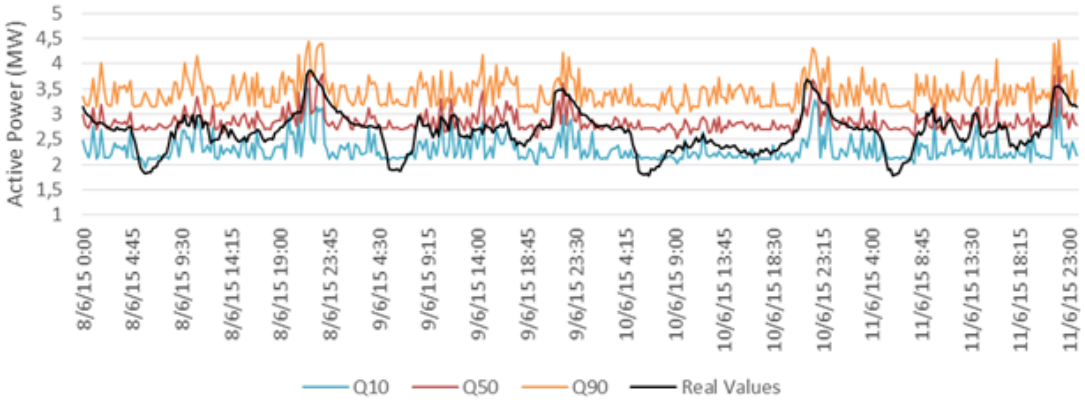

Figure 5.9: Summer Set B7 Active Power

This voltage results translate into a distribution by quantiles, as well as before, as presented in table 5.2.

Table 5.2: Quantile Distribution of B7 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
3,4%	6,4%	10,9%	10,0%	11,3%	13,8%	14,8%	11,6%	8,9%	8,7%

Figure 5.10 shows the absence of values outside the bounded interval, as well as the quantiles above and below the target frequency. Quantiles from Q_{20-30} to Q_{70-80}, are all above the 10% target frequency, quantiles Q_{0-10} and Q_{10-20} 6% and 4%, respectively, below the target frequency and quantiles Q_{80-90} and Q_{90-100} only 2% below target frequency.

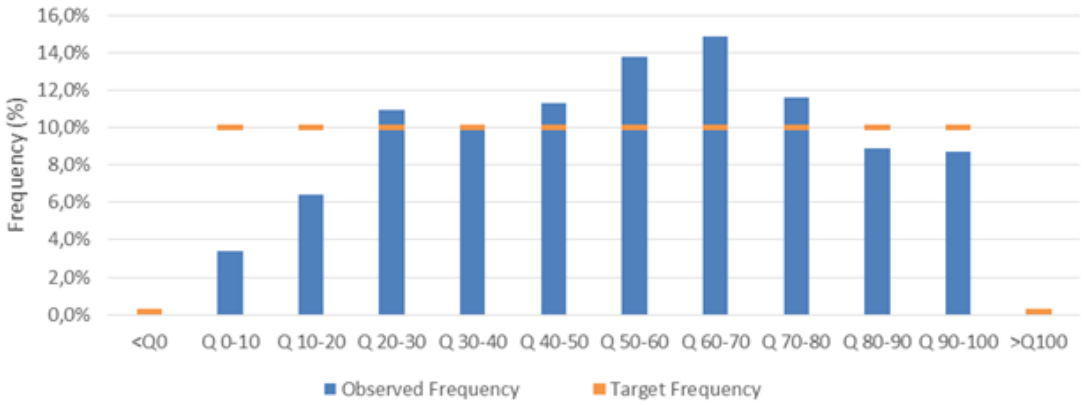

Figure 5.10: Quantile Distribution of B7 Voltage

After calculating the distribution by quantile, it was calculated the *RI* and *MAPE*, which returned a value of 74,8% and 0,90%, respectively.

5.1.3 Bus 61 - Hydro Power Generation

As stated before, B61 is characterized by being a bus with a hydro power plant. Similarly to the stated for B6, B61 voltage is highly influenced by the active power levels being produced. The voltage quantile outputs of B61 are presented in figures 5.11 and 5.12.

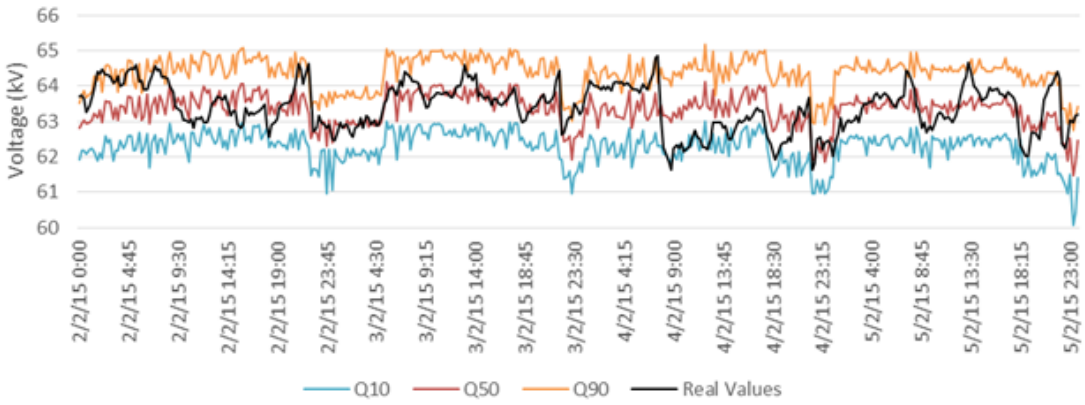

Figure 5.11: Winter Set B61 Voltage

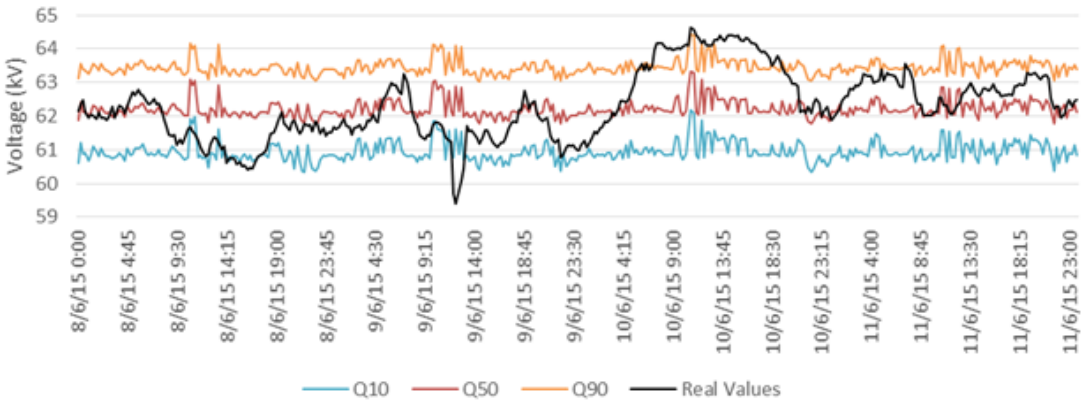

Figure 5.12: Summer Set B61 Voltage

Figures 5.11 and 5.12 also allow to take some conclusions, such as follows:

- Winter set quantile values are led significantly better by the real values;
- Winter set presents considerable voltage drops between the 2^{nd} and the 3^{rd}, the 3^{nd} and the 4^{th} and also the 4^{th} and the 5^{th} of February;
- Summer set quantiles present some regularity, contrarily to B61 voltage real values.

Voltage drops in figure 5.11 can be explained, much likely the B6 situation, by the sudden drops of active power production by the hydro power plant, evidenced in figure 5.3, as well as voltage rises displayed in figure 5.12, as they are linked to the peaks of active power production presented in figure 5.4.

The following table 5.3 presents the distribution through the quantiles for the output values.

Table 5.3: Quantile Distribution of B61 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
4,8%	8,9%	10,2%	10,7%	12,6%	12,1%	11,3%	10,9%	7,9%	10,3%

Figure 5.13 shows again the absence of values outside the bounded interval. It is important to emphasize the regularity around the target frequency, as it is observable a peak of 12% in quantiles Q_{40-50} and Q_{50-60} and a valley of 4,8% in quantile Q_{0-10}. All the other quantiles are evenly distributed.

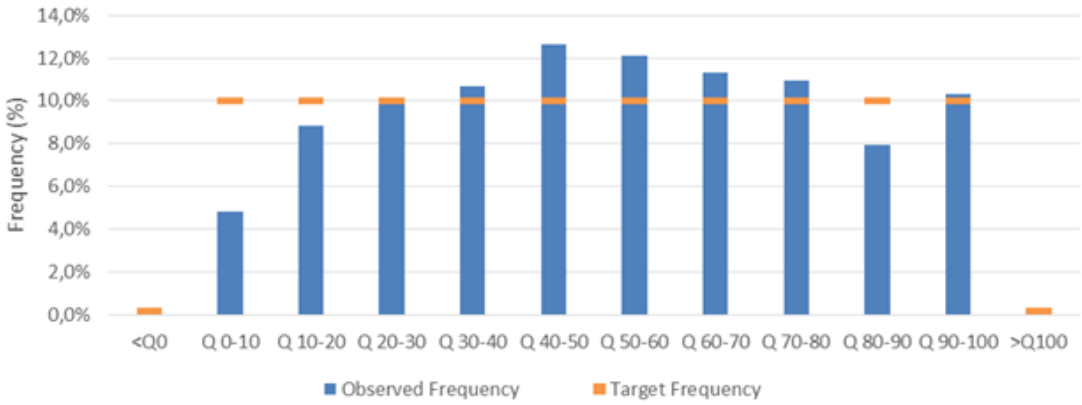

Figure 5.13: Quantile Distribution of B61 Voltage

After calculating the distribution by quantile, it was calculated the *RI* and *MAPE*, which returned a value of 83,5% and 1,02%, respectively. These results are very similar to the ones presented in section 5.1.1.

5.1.4 Bus 12 - Wind Power Generation

Just like B61, B12 is also characterized as generation bus, although with some particularities since it is a wind farm. Similarly to the stated for B6 and B61, B12 voltage is highly influenced by the active power levels being produced. The voltage quantile outputs of B12 are presented in figures 5.14 and 5.15.

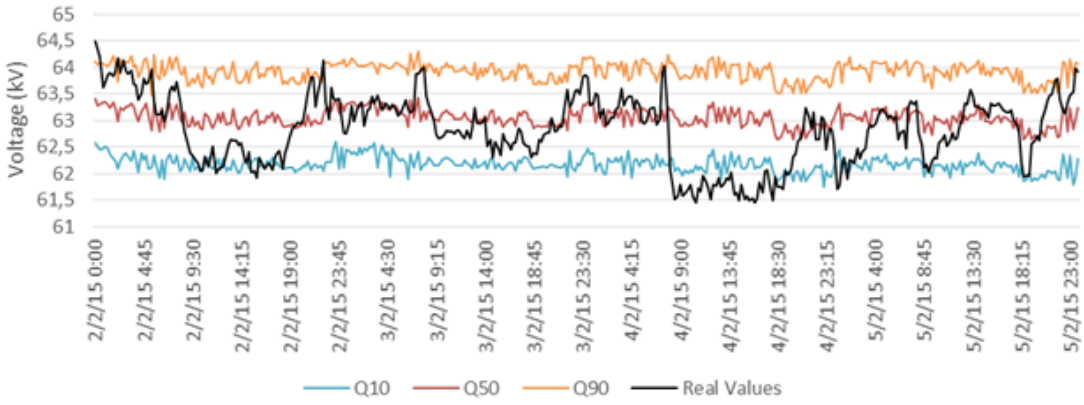

Figure 5.14: Winter Set B12 Voltage

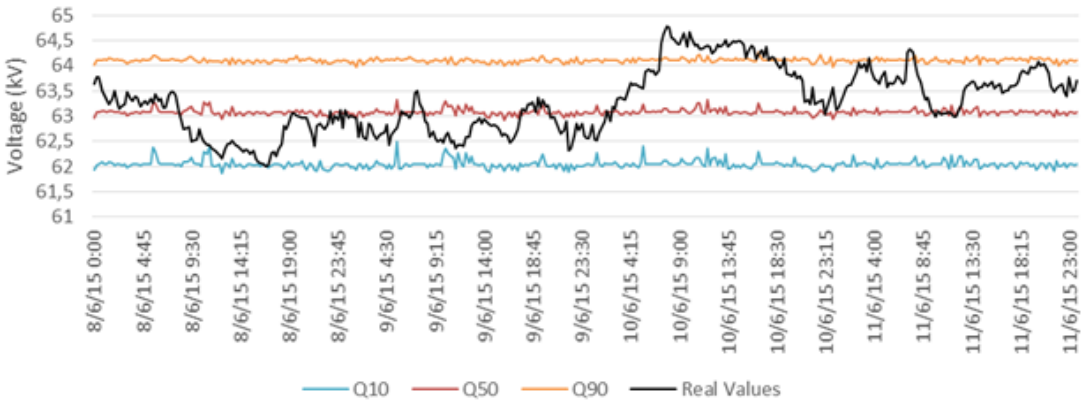

Figure 5.15: Summer Set B12 Voltage

Figures 5.14 and 5.15 allow to take some conclusions, such as follows:

- Winter set quantile values are better led by the real values than the summer set;
- Both winter and summer set quantiles present some regularity, contrarily to B61 voltage real values.

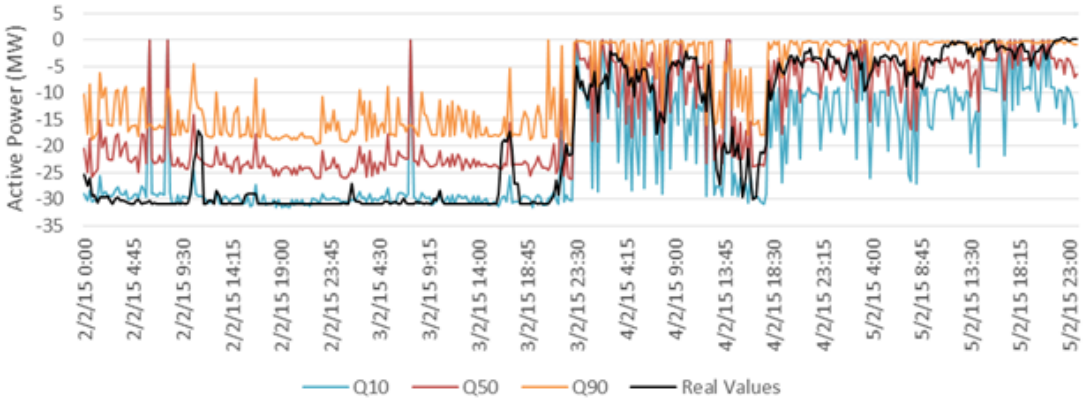

Figure 5.16: Winter Set B12 Active Power

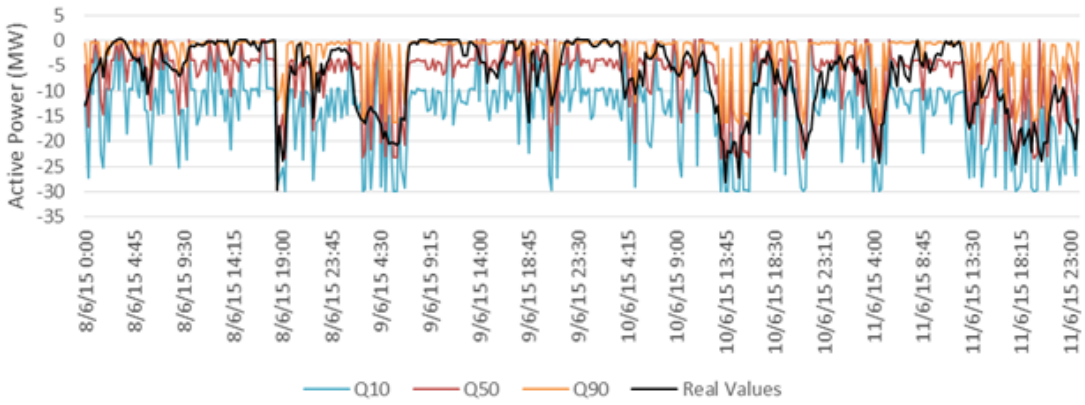

Figure 5.17: Summer Set B12 Active Power

Figures 5.16 and 5.17 elucidate very distinctively on how the output distribution quantiles are led by the input real values, on both sets. This allows to state that even though the output distribution clearly emulate the real curve, the voltage is still in independent from those values, rather being connected to other buses, such as the injection point from the transmission line, whose data could not be used since it was very fragmented.

The following table 5.4 presents the distribution through the quantiles for the output values.

Table 5.4: Quantile Distribution of B12 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
8,7%	8,6%	9,0%	11,6%	12,8%	12,6%	10,4%	11,2%	7,3%	7,7%

Figure 5.18 shows again the absence of values outside the bounded interval. It is important to emphasize the regularity around the target frequency, as it is observable a peak of 12% in quantiles Q_{40-50} and Q_{50-60} and a valley of 7,3% in quantiles Q_{80-90} and Q_{90-100}. All the other quantiles are evenly distributed. This is mainly due to the dispersed nature of the values of input active power for this bus, as it is commonly verifiable on wind power sources.

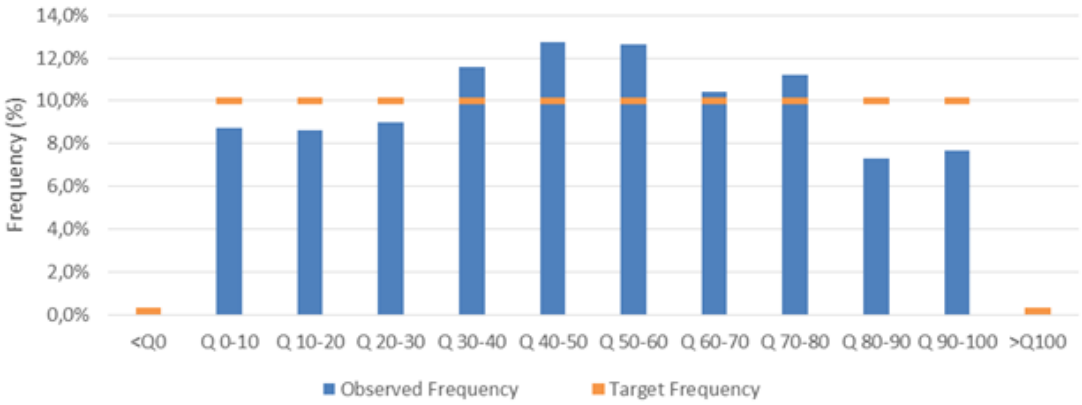

Figure 5.18: Quantile Distribution of B12 Voltage

After calculating the distribution by quantile, it was calculated the *RI* and *MAPE*, which returned a value of 82,7% and 0,83%, respectively.

5.2 Input Dispersion Reduction

In this section, the same evaluation was made similarly to the previous section, but with a tighter interval, still considering only active power inputs. The objective of this section is to compare the results to the ones presented in section 5.1. It is important to emphasize that the initial conditions were kept the same, as the only thing that was changed was the interval from 0.5 to 0.4. This values were determined resorting to a sensitivity analysis, in order to reach the minimum convergence point.

As it was explained in chapter 3, the interval, which is the target of this evaluation, is simply a boundary of the inputs, that is, instead of considering all cases regarding a given input, a first selection is made, eliminating all the cases outside those boundaries. A greater interval value will turn result in a greater range, which in turn results in a greater number of cases. This was the smallest interval value for which the method converged and returned any values, since a tighter interval would eliminate all qualified cases for certain test set instants.

5.2.1 Bus 6 - Load Bus Connected to Generation

We start our analysis by showing one of the test sets for B6 voltage output values.

One can quickly identify the differences between figure 5.19 and 5.1. The test set values obtained via smaller interval are much more responsive to smaller changes of real values, which also indicates that there are less number of qualified cases found for each test case.

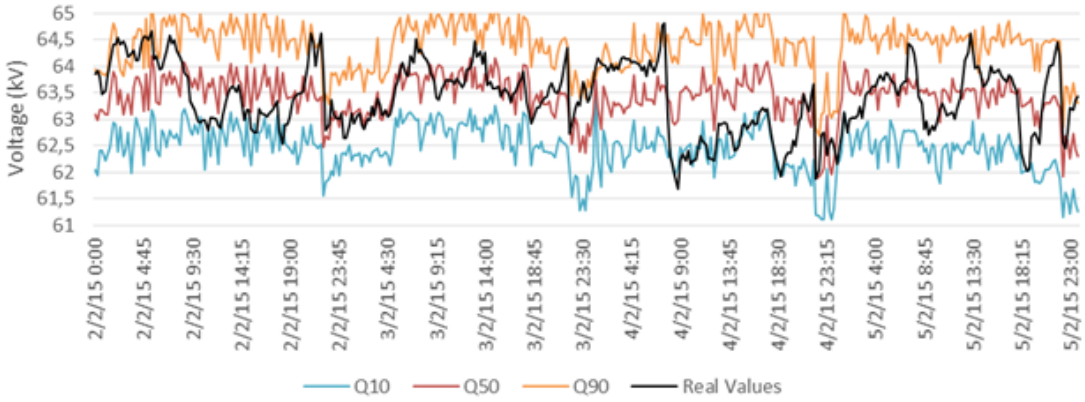

Figure 5.19: Winter Set B6 Voltage

Another indicator that alludes to better performances is the quantile distribution, presented in table 5.5 and graphically explained in figure 5.20.

Table 5.5: Quantile Distribution of B6 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
6,3%	8,1%	10,8%	11,3%	11,5%	10,8%	10,4%	11,7%	8,6%	9,5%

Although the differences are not very significant, one can see an improvement with the decrease of the interval. This is explained by a smaller number of cases which in turn eliminate cases that might have caused noise previously. By eliminating those cases, the results benefit from a finer selection, which in turn decreases the output span of values. The distribution turns out to be more disperse between the quantiles.

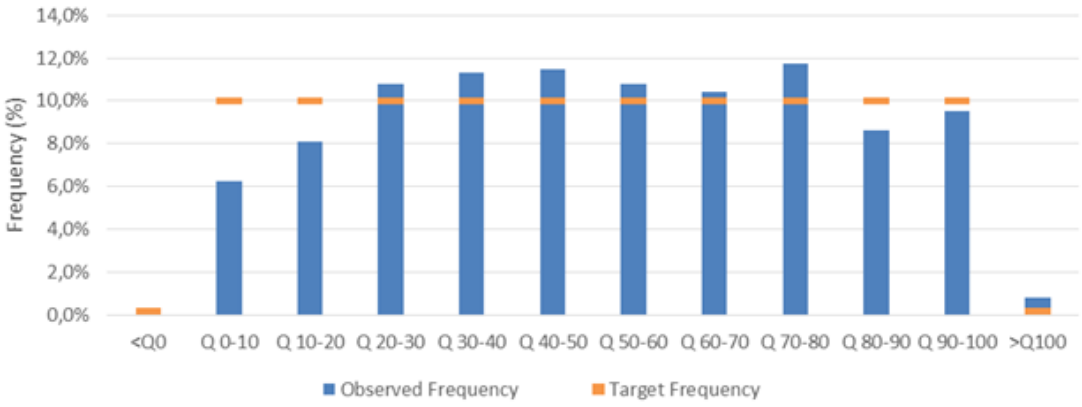

Figure 5.20: Quantile Distribution of B6 Voltage

The results shown are translated in a *RI* and *MAPE* of 85,9% and 1,00%, respectively, turning out to be more accurate, as expected. Another index was evaluated, just as for the base case, solely for B6 which contemplates the average difference between quantiles Q_{10} and Q_{90}. For this case the difference was of 2,18 kV. This difference, having 60 kV as per unit (p.u.) base voltage, represents an interval of 0.036 p.u., which deems itself acceptable.

The previous statement can be corroborated by a simple analysis of the distribution curves, from one case and another. In order to do that, a random test case was selected for the winter test set and another for the summer test set and the results are presented in tables 5.6 and 5.7.

Table 5.6: Beta Distributions Winter Set for B6 Voltage

	0.5 Interval	0.4 Interval
Alfa	2,968623	1,744528
Beta	3,260226	1,769463
Min	61,29351	62,51545
Max	65,7176	65,4612

Table 5.7: Beta Distributions Summer Set for B6 Voltage

	0.5 Interval	0.4 Interval
Alfa	6,585835	3,386069
Beta	6,567768	2,246667
Min	58,65869	59,96246
Max	65,81033	64,96479

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.21 and 5.22.

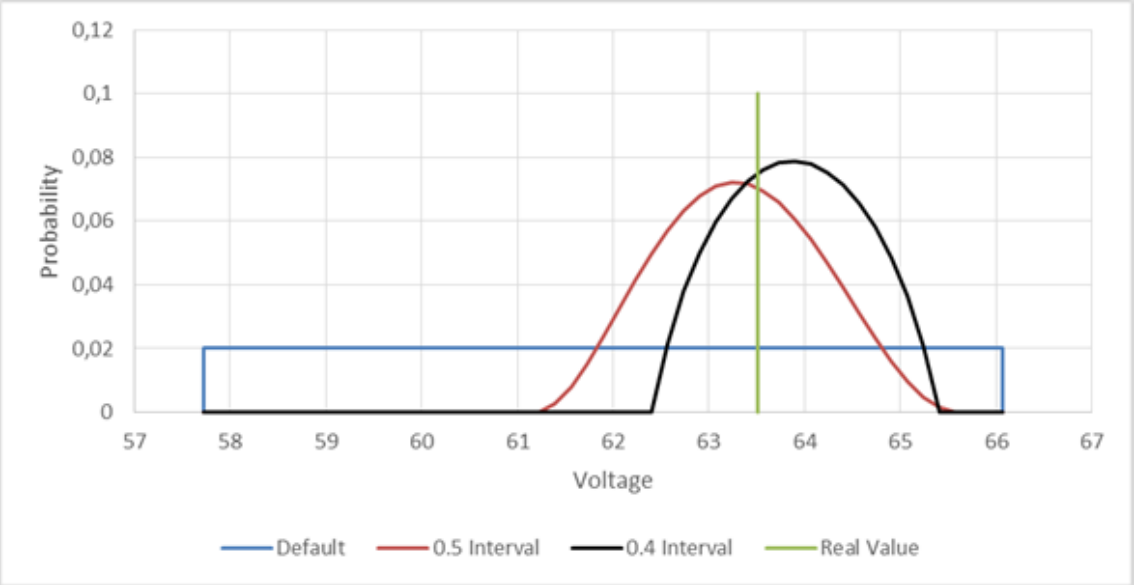

Figure 5.21: Beta Distribution of B6 Voltage, Winter Set

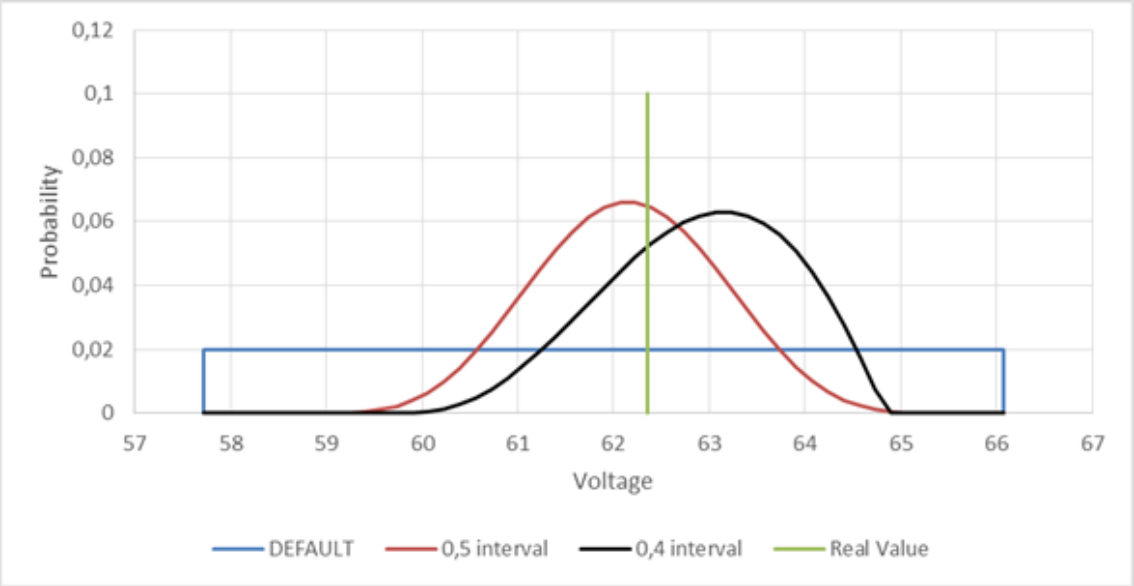

Figure 5.22: Beta Distribution of B6 Voltage, Summer Set

Although α and β are substantially higher in the base case, the fact that the interval is wider makes the curve to spread out between the maximum and minimum values, conferring to the real value a higher probability in the smaller interval winter case.

In the summer case, the benefits of a smaller intervals are not as significant, as the value of probability assigned to the real value is actually lower than the previous case. However, by analyzing table 5.7, the output interval is smaller, which helps lowering some uncertainties around the verifiable minimum and maximum values.

5.2.2 Bus 7 - Load Bus Far from Generation

The same analysis done for B6 will be done to B7. Again, one can easily identify the differences between figure 5.23 and 5.6. The test set values obtained via smaller interval are much more responsive to smaller changes of real values, which also indicates that there are less number of qualified cases in the training set found for each test case.

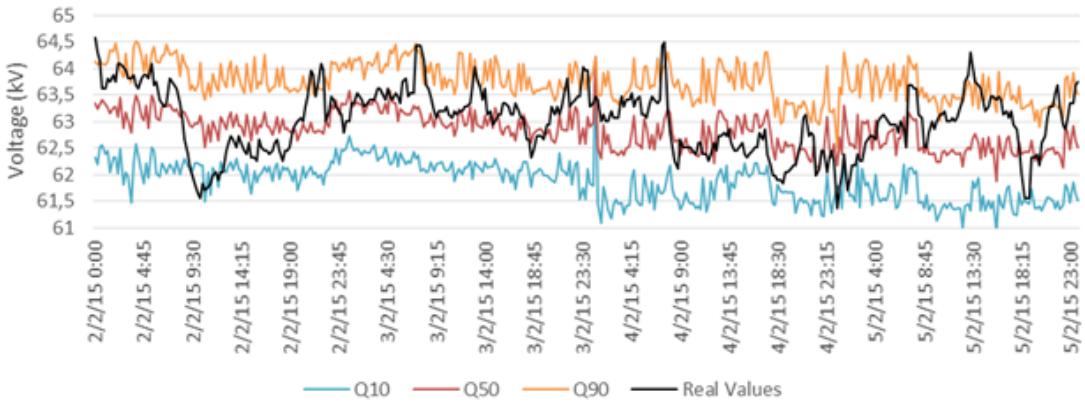

Figure 5.23: Winter Set B7 Voltage

Another indicator that alludes to better performances is the quantile distribution, presented in table 5.8 and graphically explained in figure 5.24.

Table 5.8: Quantile Distribution of B7 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
3,3%	6,3%	9,8%	11,6%	8,7%	13,7%	12,9%	13,0%	9,2%	10,9%

The differences are not as significant as in the previous case, but one can see an improvement with the decrease of the interval. This is again explained by a smaller number of cases which in turn eliminate cases that might have caused noise previously. The quantile distribution is presented as follows.

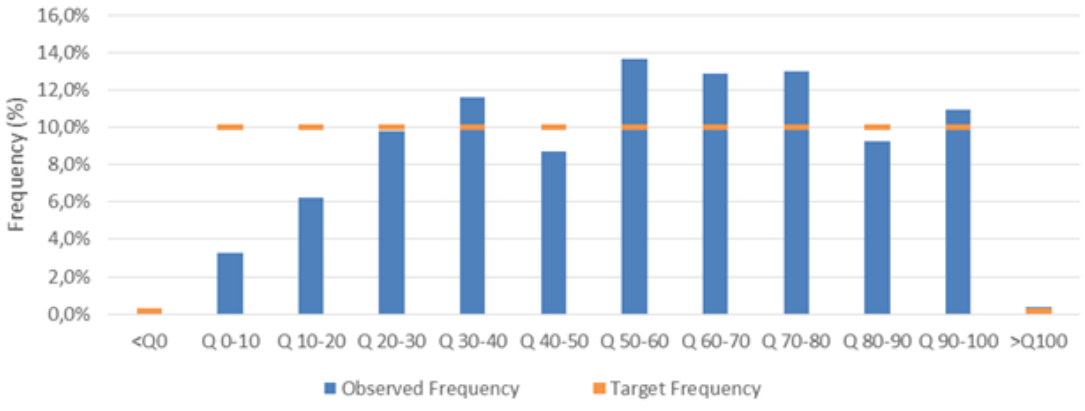

Figure 5.24: Quantile Distribution of B7 Voltage

The results shown are translated in a *RI* and *MAPE* of 75,1% and 0,92%, respectively, turning out to be more accurate, again as expected.

Then it was selected the same test cases as before, one for the winter test set and another for the summer test set, as the results are presented in tables 5.9 and 5.10.

Table 5.9: Beta Distributions Winter Set for B7 Voltage

	0.5 Interval	0.4 Interval
Alfa	2,24715	2,111688
Beta	2,00407	2,680263
Min	61,15909	61,77273
Max	64,29546	64,9091

Table 5.10: Beta Distributions Summer Set for B7 Voltage

	0.5 Interval	0.4 Interval
Alfa	6,41543	2,925178
Beta	6,04383	2,446551
Min	59,18182	60,40909
Max	65,93182	64,97727

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.25 and 5.26.

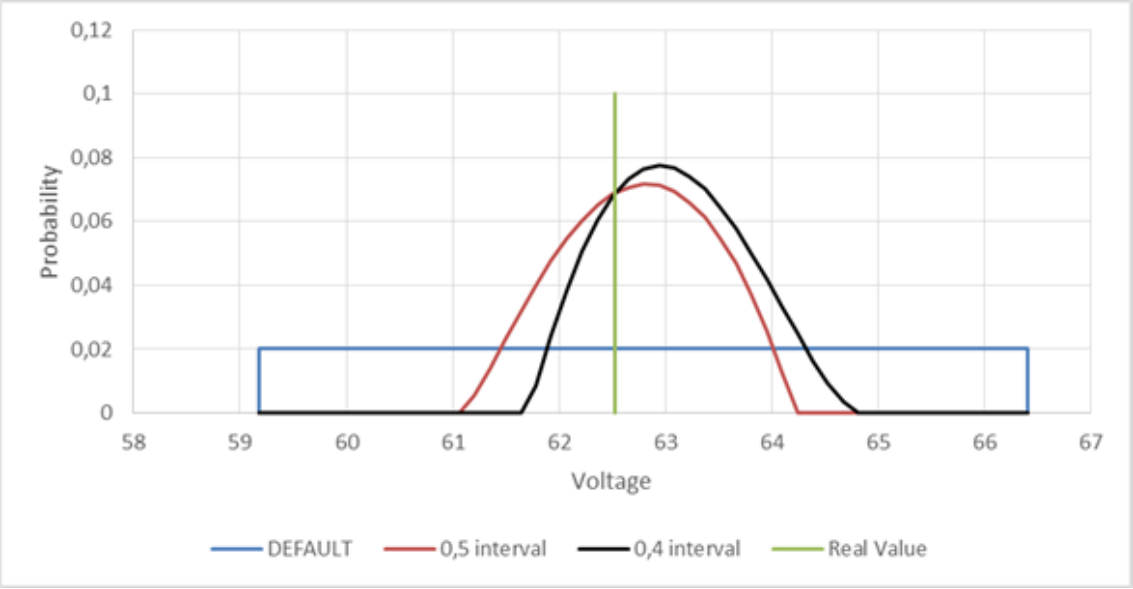

Figure 5.25: Beta Distribution of B7 Voltage, Winter Set

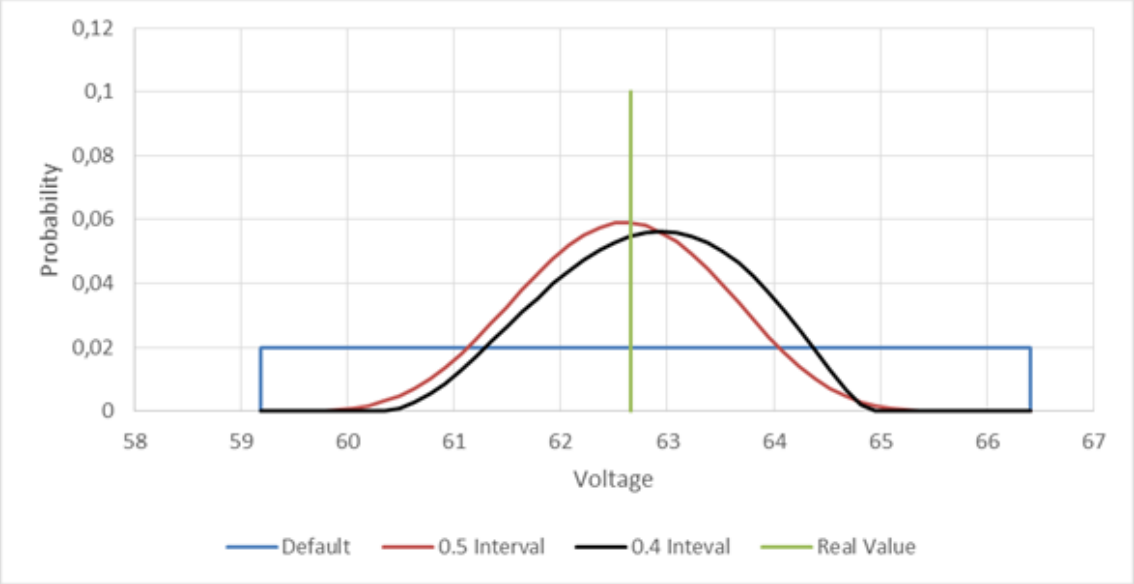

Figure 5.26: Beta Distribution of B7 Voltage, Summer Set

In the winter case, although the probability assigned to the real value is the same in both cases, it is possible to verify that there was a shift to the right, accounting for higher values of voltage. This could be explained by the random function regarding the determination the interval maximum and minimum limits. Even though the selected interval is smaller, the limits can change, not only the amplitude, but also the position of the test value in the interval, which can qualify cases totally different from the base case.

In the summer case, there is a slight variation of the minimum and maximum value, accompanied with an equally slight shift of the curve to the right.

5.2.3 Bus 61 - Hydro Power Generation

The test set values obtained via smaller interval, figure 5.27, are again more responsive to smaller changes of real values than the base case, figure 5.11.

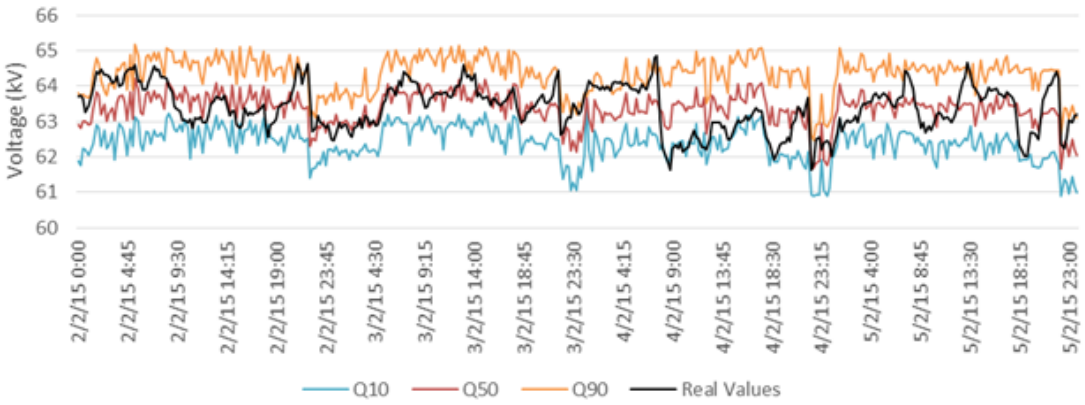

Figure 5.27: Winter Set B61 Voltage

Another indicator that alludes to better performances is the quantile distribution, presented in table 5.11 and graphically explained in figure 5.28.

Table 5.11: Quantile Distribution of B61 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
5,3%	7,2%	11,1%	11,6%	11,1%	11,1%	10,5%	11,5%	9,6%	10,0%

The differences are relevant, as one can see an improvement with the decrease of the interval. The quantile distribution is presented as follows.

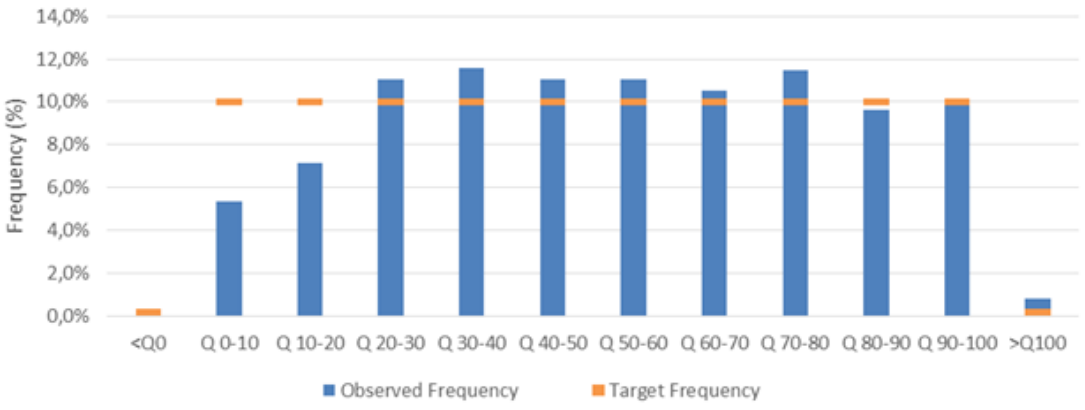

Figure 5.28: Quantile Distribution of B61 Voltage

The results shown are translated in a *RI* and *MAPE* of 85,3% and 1,01%, respectively, turning out to be more accurate, again as expected.

Then it was selected the same test cases as before, one for the winter test set and another for the summer test set, as the results are presented in tables 5.12 and 5.13.

Table 5.12: Beta Distributions Winter Set for B61 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	2,968623	3,754161
Beta	3,260226	4,024068
Min	61,29351	61,44625
Max	65,7176	65,65759

Table 5.13: Beta Distributions Summer Set for B61 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	6,585835	4,24563
Beta	6,567768	3,821228
Min	58,65869	58,6096
Max	65,81033	64,52838

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.29 and 5.30.

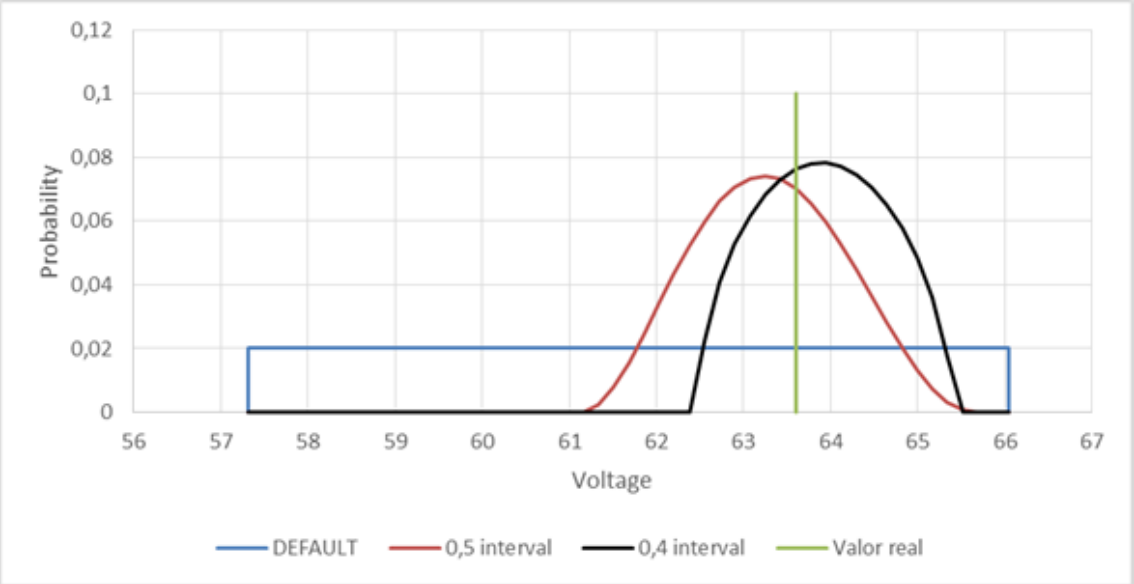

Figure 5.29: Beta Distribution of B61 Voltage, Winter Set

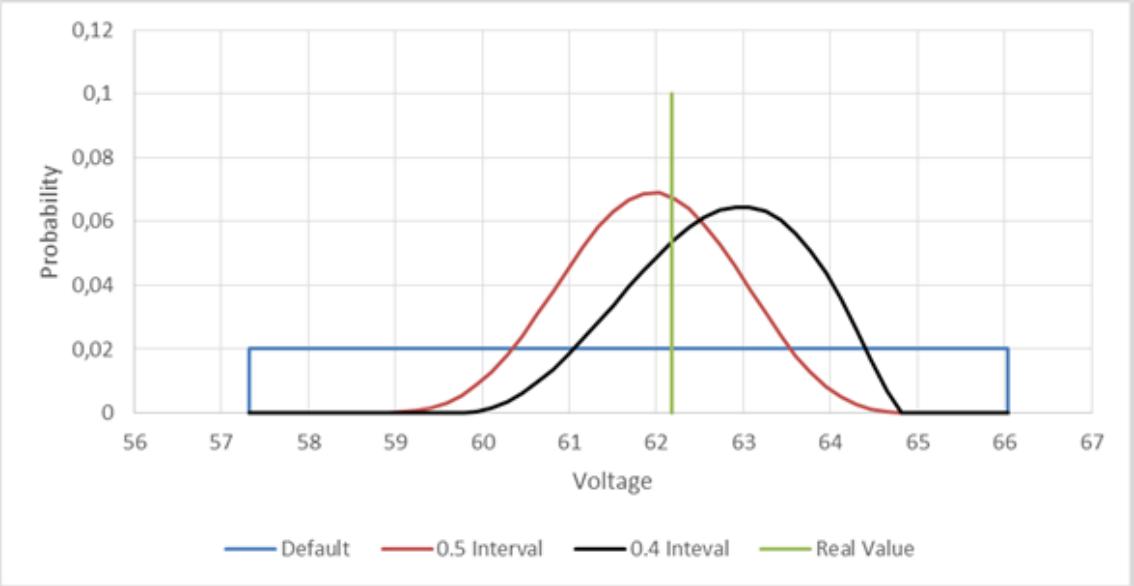

Figure 5.30: Beta Distribution of B61 Voltage, Summer Set

This case is very similar to the B6 case, mainly because they are connected and, therefore, influence each other. Again, in the winter test case, although the probability assigned to the real value is almost the same in both cases, it is possible to verify that there was a shift to the right, discarding lower values of voltage incorporated in the previous interval. In the summer test case, although the value of probability assigned to the real value is lower than the previous case, most of the lower values of the curve also have a lower probability, as well as increased minimum and decreased maximum values.

5.2.4 Bus 12 - Wind Power Generation

Finally, to conclude the study of the smaller interval, it is presented the results for B12. The test set values obtained via smaller interval, figure 5.31, are again more responsive to smaller changes of real values than the base case, figure 5.14.

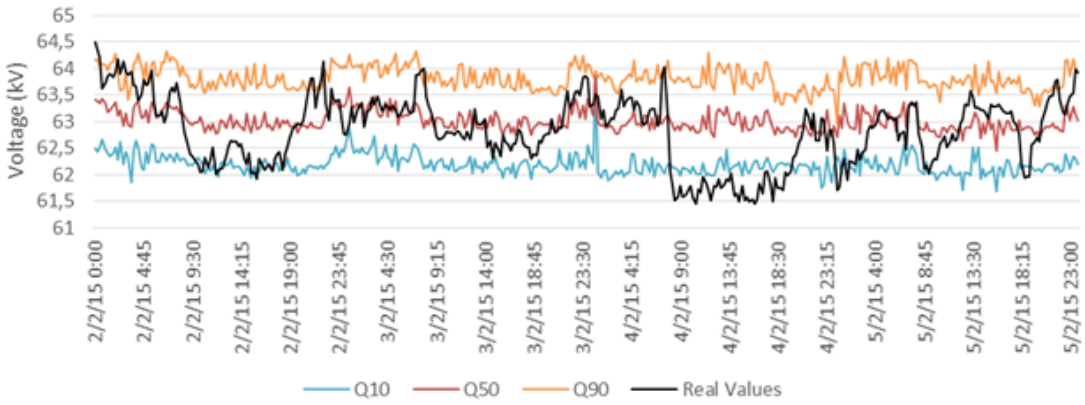

Figure 5.31: Winter Set B12 Voltage

Another indicator that alludes to better performances is the quantile distribution, presented in table 5.14 and graphically explained in figure 5.32.

Table 5.14: Quantile Distribution of B12 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
8,7%	7,0%	9,0%	9,4%	13,7%	12,0%	10,8%	11,8%	7,6%	8,7%

The differences are not so significant, as one can see some improvement with the decrease of the interval, namely on central quantiles, as it is presented next.

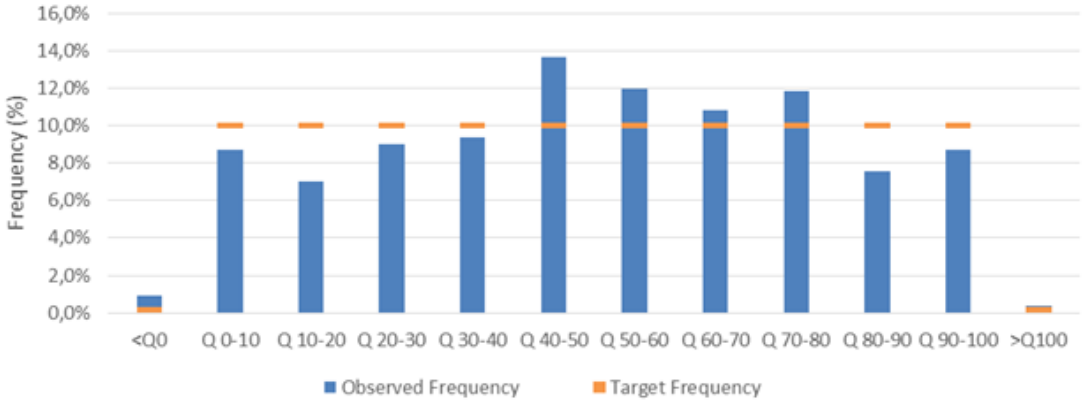

Figure 5.32: Quantile Distribution of B61 Voltage

The results shown are translated in a *RI* and *MAPE* of 82,1% and 0,82%, respectively, turning out to be equally accurate.

Then it was selected the same test cases as before, one for the winter test set and another for the summer test set, as the results are presented in tables 5.15 and 5.16.

Table 5.15: Beta Distributions Winter Set for B12 Voltage

	0.5 Interval	0.4 Interval
Alfa	2,286982	2,300165
Beta	2,364291	2,270871
Min	61,46869	61,49341
Max	64,41211	64,55219

Table 5.16: Beta Distributions Summer Set for B12 Voltage

	0.5 Interval	0.4 Interval
Alfa	5,461713	3,645097
Beta	5,621395	2,959827
Min	60,3609	60,85162
Max	65,80371	65,0658

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.33 and 5.34.

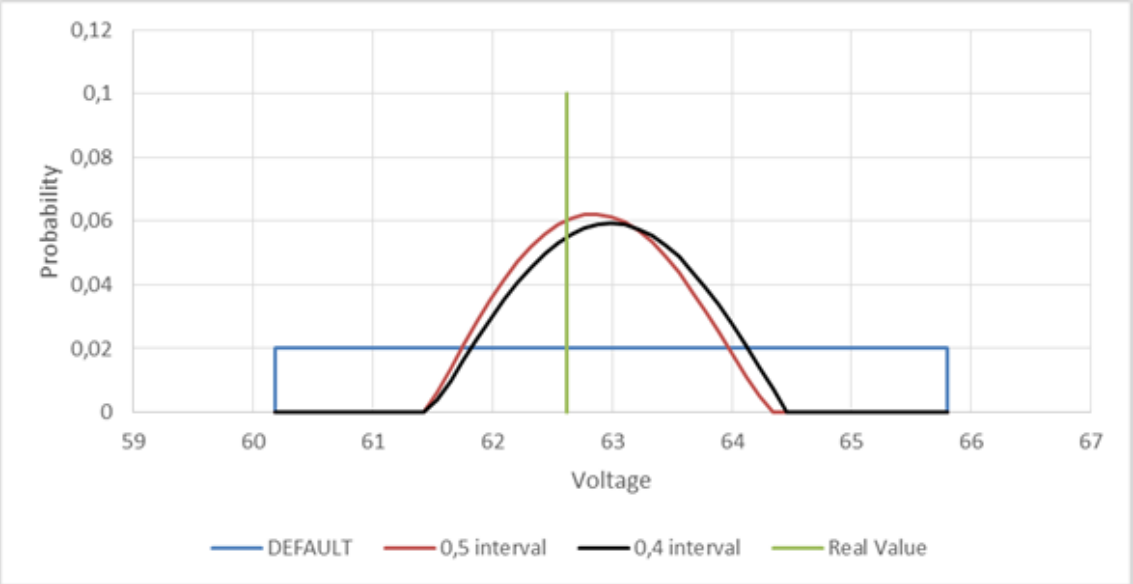

Figure 5.33: Beta Distribution of B12 Voltage, Winter Set

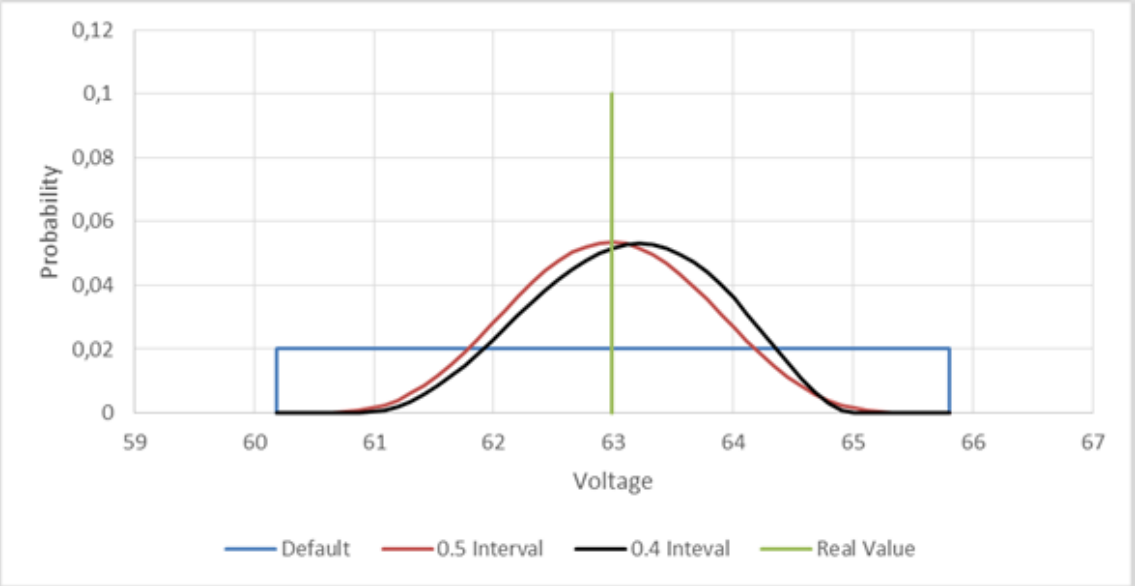

Figure 5.34: Beta Distribution of B12 Voltage, Summer Set

The similarity between the results presented in both winter and summer sets is a clear evidence of how well distributed are the inputs of this bus. By changing the interval, the differences between both cases turn out to be almost insignificant. This could be explained by the fact that this bus is far away from the buses at study and that could mean that by varying the information of the other buses, the output of this bus will not be as influenced as the others. Another explanation to this is that, as this bus is connected to the injection point from the transmission grid, the voltage on B12 can be rather influenced by the voltage of the said injection point. In both cases, there are little changes to the minimum and maximum values, being possible to observe a slight shift of the curve towards higher voltage values.

5.3 Inclusion of New Input Variables

In this section it will be studied the scenario where not only active power is available, but reactive power is also available. Since this is a high voltage grid, reactive power can really have an impact since the inductance is higher than the resistance and, therefore, voltage components (magnitude and angle) can be decoupled, so active power would have a higher influence in the voltage angle and reactive power in the voltage magnitude. The results shown are compared to the base case, in section 5.1, since the interval is maintained at 0.5, changing only the amount of inputs to contemplate reactive power alongside active power.

5.3.1 Bus 6 - Load Bus Connected to Generation

Again, it is notorious the difference between figures 5.35 and 5.1. The test set values obtained with reactive power are much more responsive to smaller changes of real values, which also indicates that there are less number of qualified cases found for each test case. This is due to a tighter selection of cases when considering reactive power inputs.

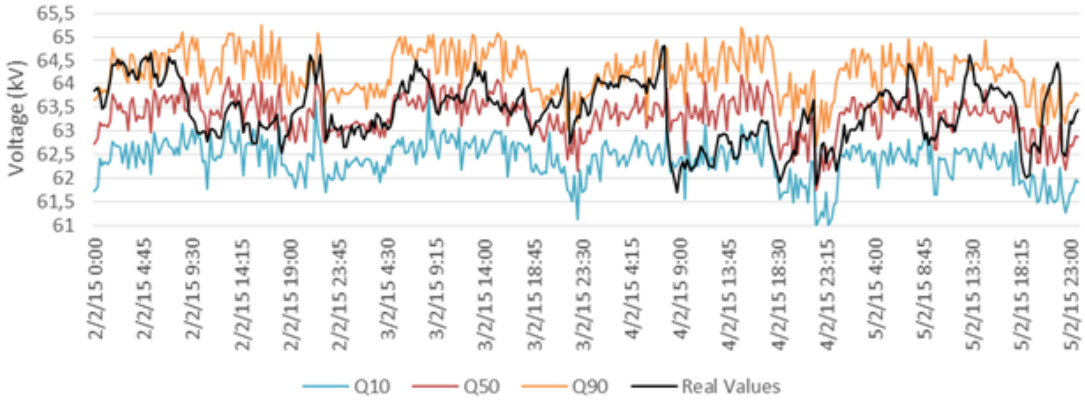

Figure 5.35: Winter Set B6 Voltage

Figures 5.36 and 5.37 show how the input is led by the real value, being translated in a good approximation for both summer and winter set. This means that if there is a correlation between reactive power and voltage in this bus, output voltage should have better indexes.

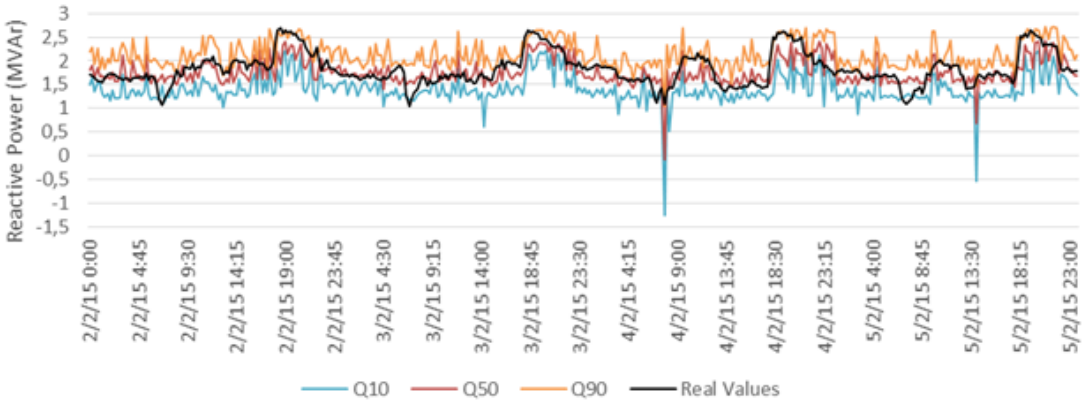

Figure 5.36: Quantile Distribution of B6 Reactive Power, Winter Set

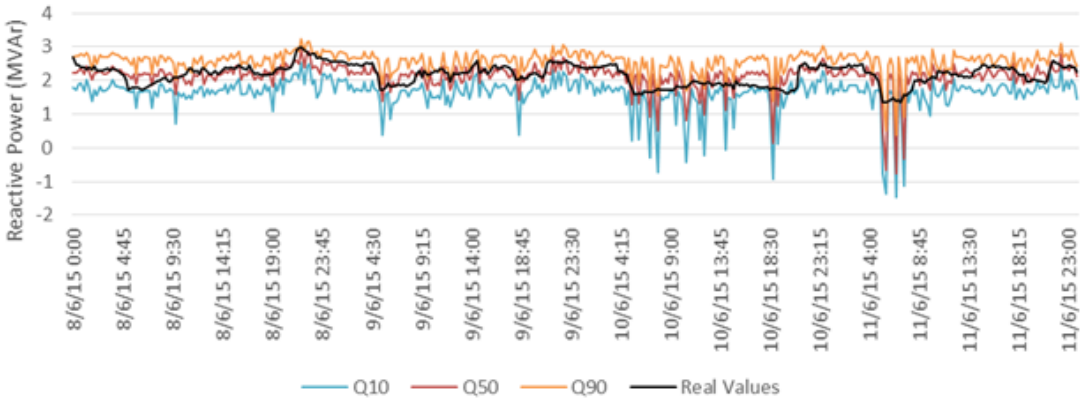

Figure 5.37: Quantile Distribution of B6 Reactive Power, Summer Set

However, one must look to the quantile distribution, presented in table 5.17 and graphically explained in figure 5.38, to better understand if the tighter selection really indicates better performances.

Table 5.17: Quantile Distribution of B6 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
4,6%	7,4%	7,6%	9,8%	9,6%	12,4%	11,3%	13,8%	10,7%	12,0%

With the information presented before, one can assess that there was no improvement in this case by adding the reactive power inputs. This is explained by a smaller number of cases which in turn may have eliminated cases that might have been a better choice to the output value. The distribution turns out to be more concentrated in the center and rightmost quantiles.

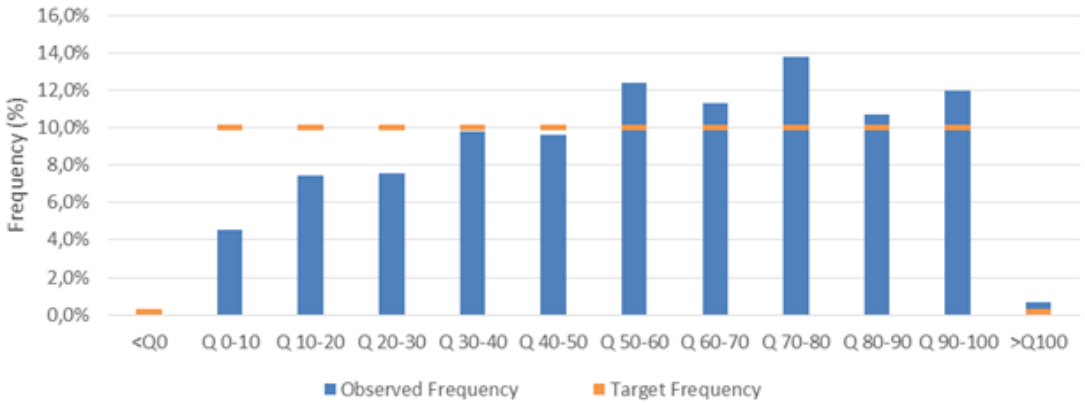

Figure 5.38: Quantile Distribution of B6 Voltage

The results shown are translated in a *RI* and *MAPE* of 78,8% and 1,03%, respectively, turning out to be less accurate. Another index was evaluated, just as for the base case, solely for B6 which contemplates the average difference between quantiles Q_{10} and Q_{90}. Contrarily to the results present before, for this case the difference was of 2,17 kV. This difference, having 60 kV as per unit (p.u.) base voltage, represents an interval of 0.036 p.u., which deems itself acceptable.

The previous statement can be corroborated by a simple analysis of the distribution curves, from one case and another. In order to do that, a random test case was selected for the winter test set and another for the summer test set and the results are presented in tables 5.18 and 5.19.

Table 5.18: Beta Distributions Winter Set for B6 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	2,968623	1,744528
Beta	3,260226	1,769463
Min	61,29351	62,51545
Max	65,7176	65,4612

Table 5.19: Beta Distributions Summer Set for B6 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	6,585835	3,386069
Beta	6,567768	2,246667
Min	58,65869	59,96246
Max	65,81033	64,96479

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.39 and 5.40.

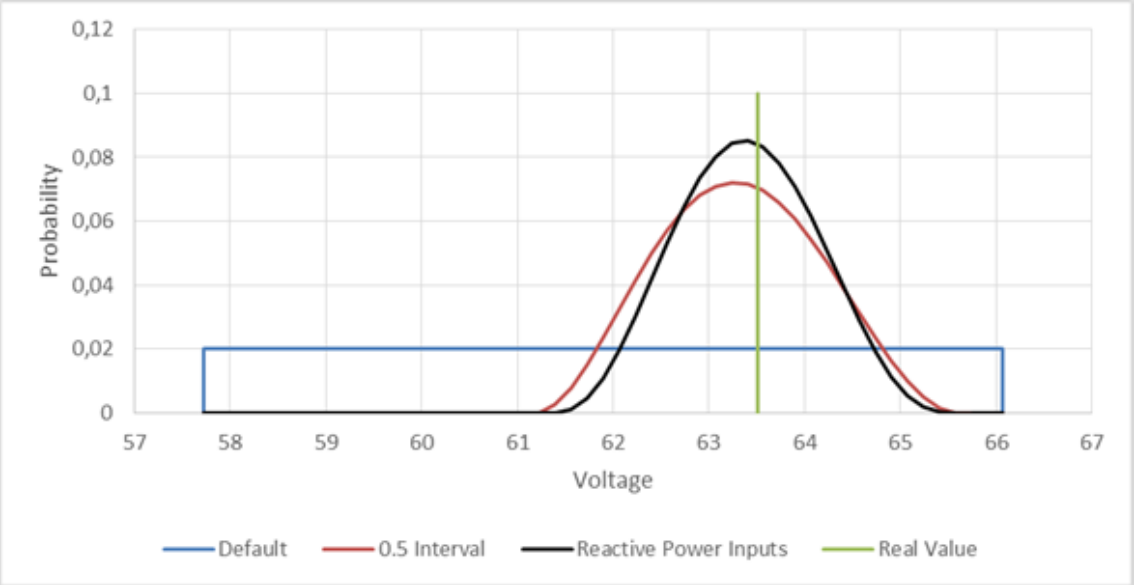

Figure 5.39: Beta Distribution of B6 Voltage, Winter Set

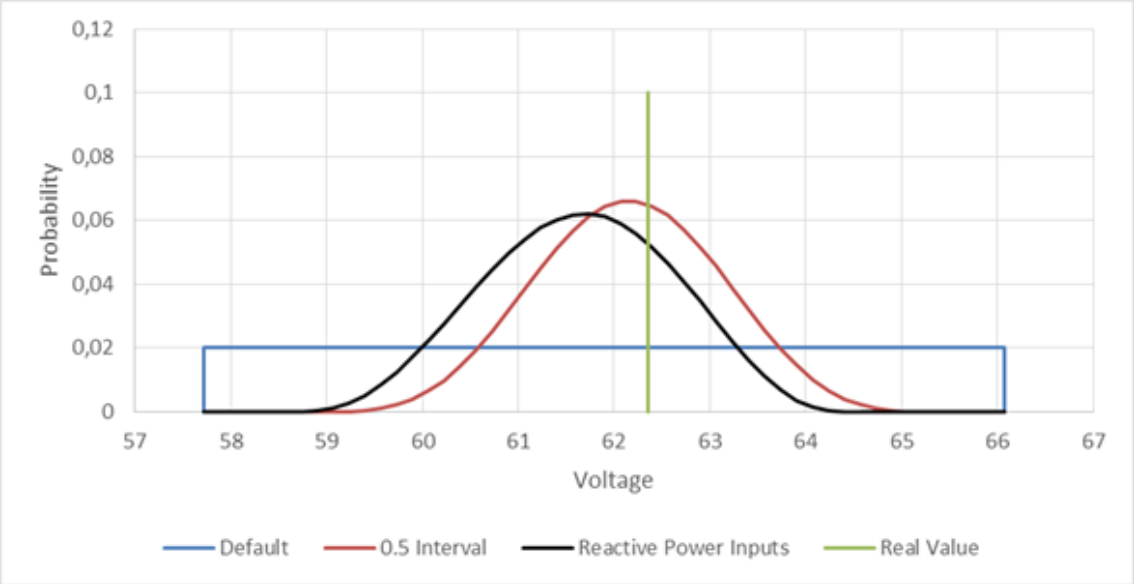

Figure 5.40: Beta Distribution of B6 Voltage, Summer Set

There are two different evaluations that can be made here, one for the winter set and one for the summer set. Starting by the winter set, the results clearly show a major improvement over the base case. Not only the interval is smaller, as evidenced in table 5.18, but also the real value turns out to have a higher probability of occurrence. This means that in this case the program correctly identified the higher activation values to shrink the interval and, consequently, to tighten up the curve.

On the other hand, the summer set does not present improvement, as the interval amplitude is almost the same. Also the curve moved away from the real value, lowering the probability of occurrence.

One can infer from the information presented in this case that reactive power inputs can be useful, but just in certain situations. Again, in this case, reactive power inputs presented themselves as useful in the winter set, but not as much in the summer set.

5.3.2 Bus 7 - Load Bus Far from Generation

There is not much difference between figure 5.41 and 5.6, as B7 is a bus that only contemplates loads, both active and reactive.

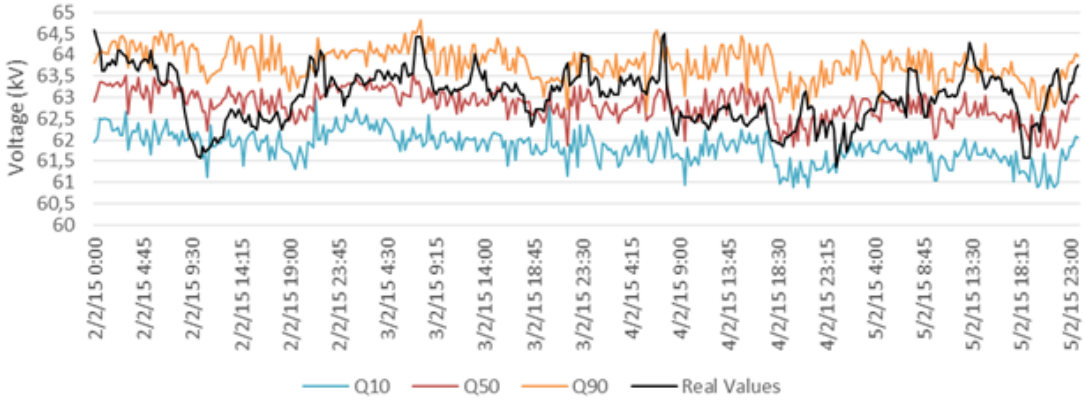

Figure 5.41: Winter Set B7 Voltage

To better evaluate the performance, we resort to the quantile distribution, presented in table 5.20 and graphically explained in figure 5.42.

Table 5.20: Quantile Distribution of B7 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
2,1%	6,3%	7,8%	10,2%	10,3%	13,5%	13,3%	16,0%	8,7%	11,1%

The results are not significantly worse, but it is clear, by comparing figures 5.42 and 5.10, that the results are much more approximated to the center of the distribution, which indicates a wider curve.

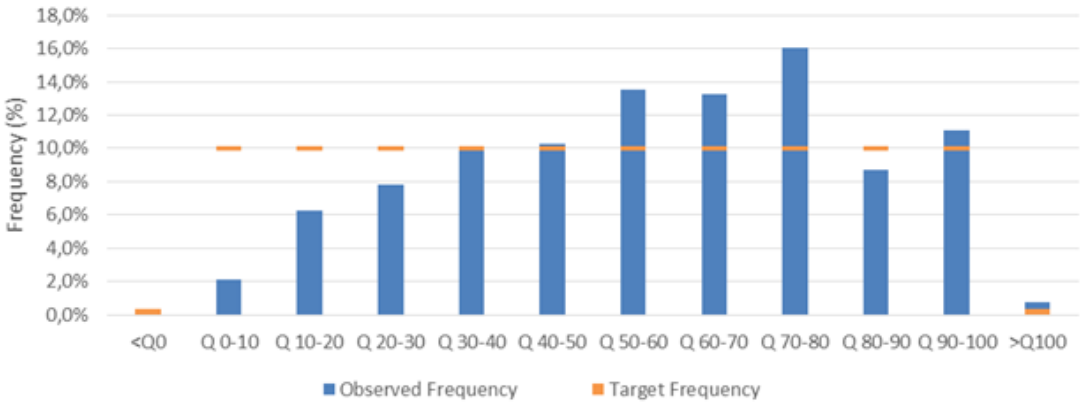

Figure 5.42: Quantile Distribution of B7 Voltage

However, the results are still satisfying, as it was calculated a *RI* and *MAPE* of 70,5% and 0,93%, respectively, turning out to be a bit less accurate.

Then it was selected the same test cases as before, one for the winter test set and another for the summer test set, as the results are presented in tables 5.21 and 5.22.

Table 5.21: Beta Distributions Winter Set for B7 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	2,24715	5,012773
Beta	2,00407	4,498286
Min	61,15909	60,40909
Max	64,29546	65,04546

Table 5.22: Beta Distributions Summer Set for B7 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	6,41543	4,810316
Beta	6,04383	4,873067
Min	59,18182	59,18182
Max	65,93182	65,52273

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.43 and 5.44.

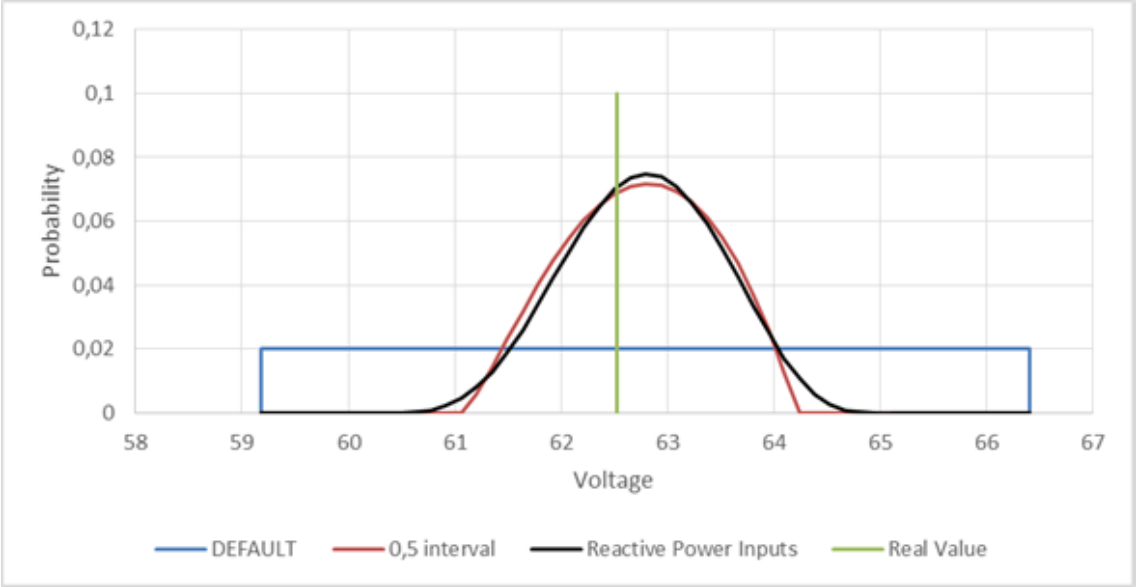

Figure 5.43: Beta Distribution of B7 Voltage, Winter Set

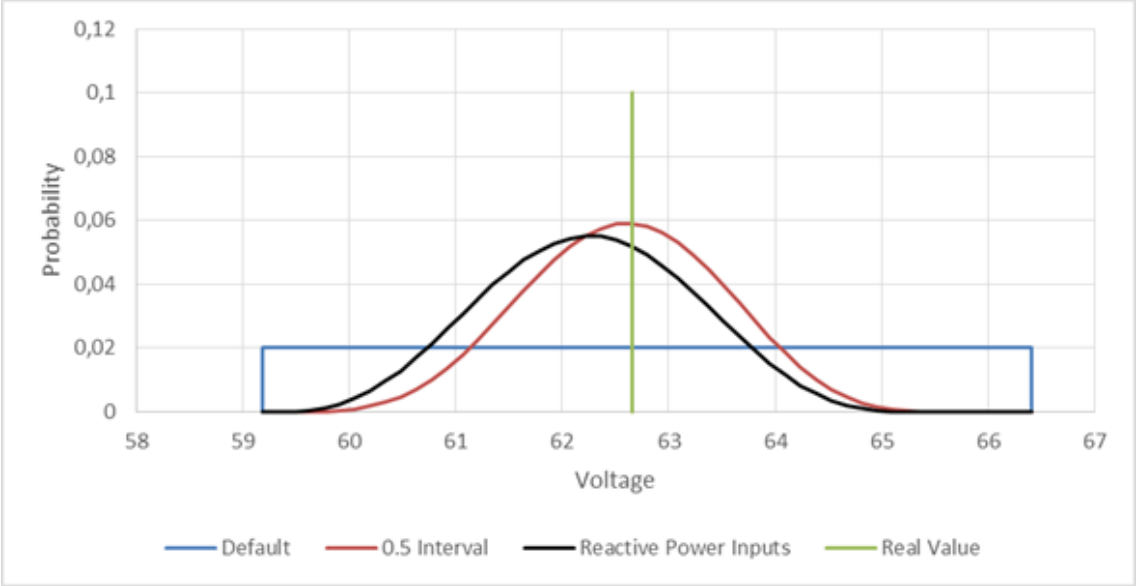

Figure 5.44: Beta Distribution of B7 Voltage, Summer Set

Figures 5.25 and 5.26 can confirm previous statements telling that the results are not significantly worse. As one can observe, for the selected case, in the winter set, although the interval is larger than the base case, α and β values are higher, which in turn shrink the curve. There is not much improvement but, again, the results are still satisfactory, given the aperture of the interval used. In the summer set we experience a slight variation of the curve to the left, indicating that for the qualified cases the voltage is lower than the real value.

5.3.3 Bus 61 - Hydro Power Generation

The test set values obtained resorting to reactive power inputs, figure 5.45, are again more responsive to smaller changes of real values than the base case, figure 5.11.

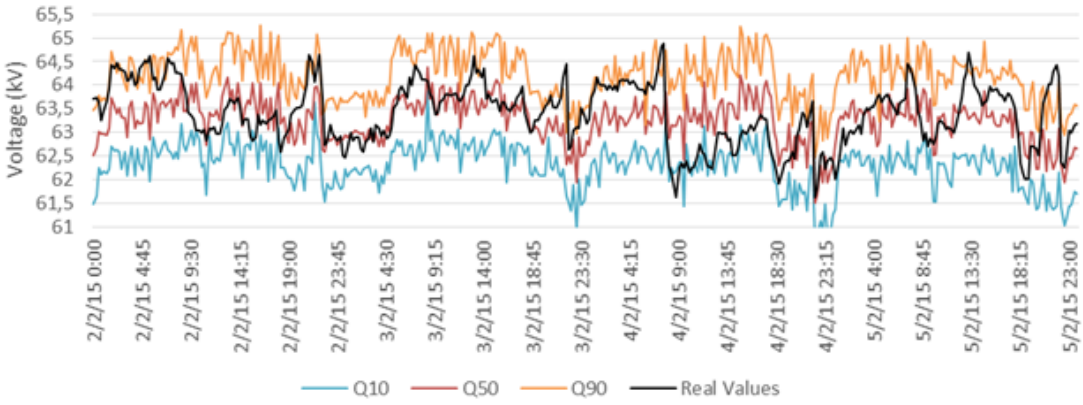

Figure 5.45: Winter Set B61 Voltage

Once again we resort to the quantile distribution, presented in table 5.23, being graphically explained in figure 5.46.

Table 5.23: Quantile Distribution of B61 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
4,2%	6,4%	8,2%	7,9%	10,4%	12,2%	12,8%	12,8%	11,2%	12,5%

The differences between this and the base case are relevant, as one can see observe a shift from the left quantiles to the center. The quantile distribution is presented as follows.

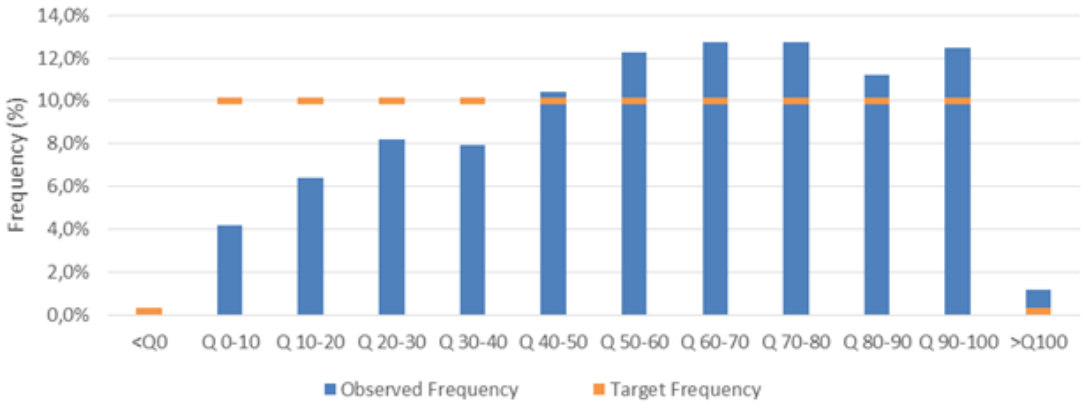

Figure 5.46: Quantile Distribution of B61 Voltage

The results shown are translated in a *RI* and *MAPE* of 74,8% and 1,05%, respectively, turning out to be again less accurate.

Then it was selected the same test cases as before, one for the winter test set and another for the summer test set, as the results are presented in tables 5.24 and 5.25.

Table 5.24: Beta Distributions Winter Set for B61 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	2,983113	3,636454
Beta	3,33614	3,975378
Min	61,24232	61,44073
Max	65,7774	65,71126

Table 5.25: Beta Distributions Summer Set for B61 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	6,41543	4,810316
Beta	6,04383	4,873067
Min	59,18182	59,18182
Max	65,93182	65,52273

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.47 and 5.48.

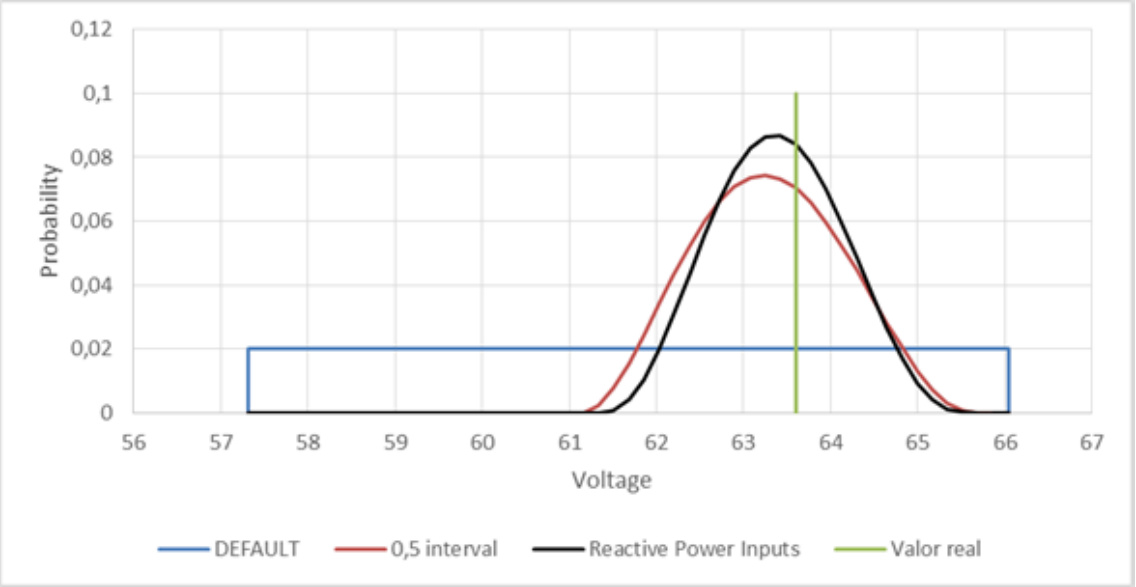

Figure 5.47: Beta Distribution of B61 Voltage, Winter Set

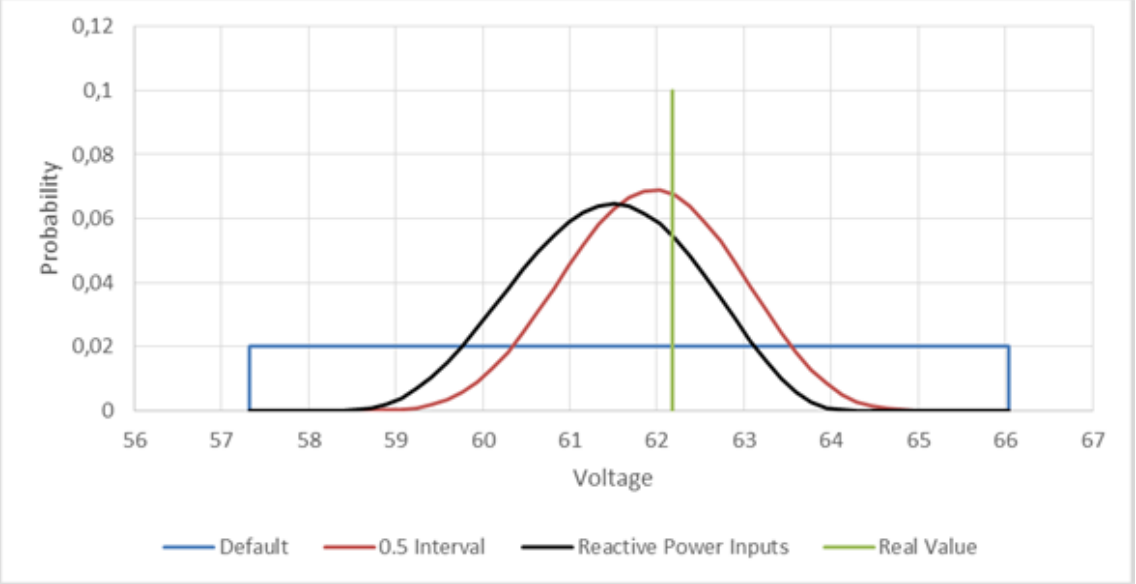

Figure 5.48: Beta Distribution of B61 Voltage, Summer Set

This case is very similar to the previous one as the winter set benefits from the addition of reactive power inputs. This is explained by the capacity of hydro power plants of setting up reactive power output control in order to directly influence the voltage in high voltage grids. Very much like the B6 analysis, and being the two buses connected, once again our method can correctly identify the reactive power input in B61, as figure 5.49 shows.

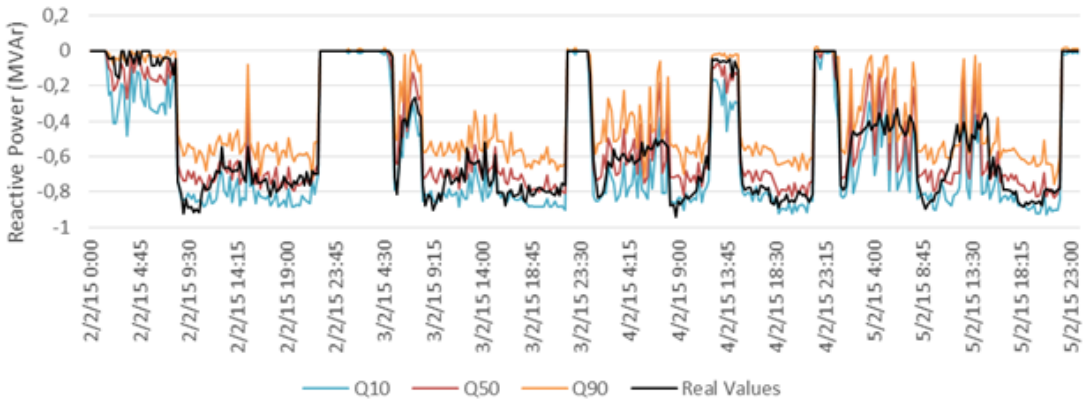

Figure 5.49: Quantile Distribution of B61 Reactive Power, Winter Set

Although being less significant than active power inputs, some of the voltage drops can be explained by the variations in figure 5.49. This results in a better approximation, as it is shown in figure 5.24. The interval is tighter than the base case and the curve itself shrinks, to give the real value higher values of probability.

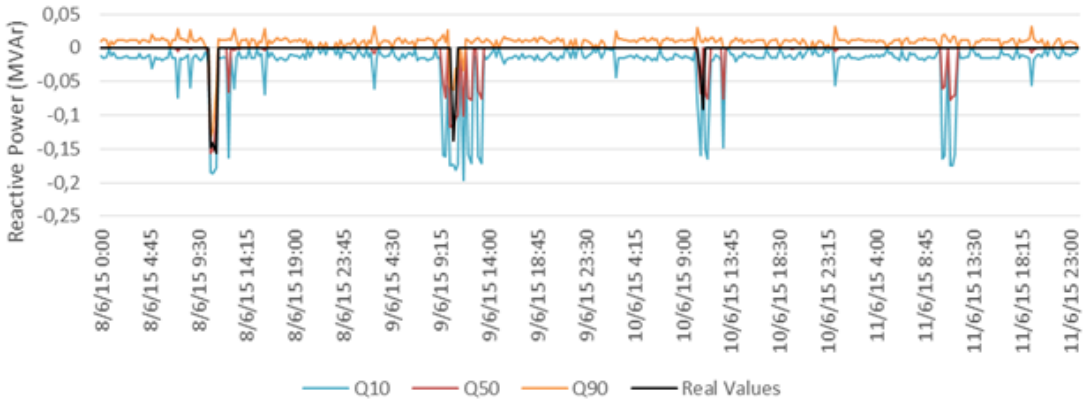

Figure 5.50: Quantile Distribution of B61 Reactive Power, Summer Set

On the other hand, there is the summer set. The results are not as satisfactory as the winter set's, very much because of the reactive power inputs of the set. As one can observe in figure 5.50, the values for the summer test set are very close to zero, within rare exceptions. This can explain the behaviour of the curve presented in figure 5.25, since there is little to no influence in

the voltage output. This is a case where is prejudicial to use reactive power inputs in the process of determination of voltage outputs.

5.3.4 Bus 12 - Wind Power Generation

In order to conclude the study of the reactive power inputs, it is presented the results for B12. The test set values obtained using those set of inputs, figure 5.51, are again more responsive to smaller changes of real values than the base case, figure 5.14.

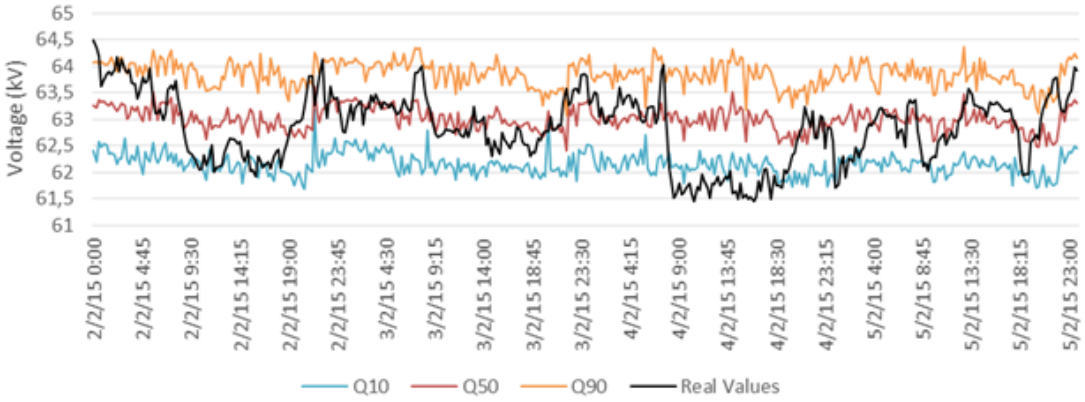

Figure 5.51: Winter Set B12 Voltage

Another indicator that alludes to better performances is the quantile distribution, presented in table 5.26 and graphically explained in figure 5.52.

Table 5.26: Quantile Distribution of B12 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
7,9%	6,6%	7,9%	10,3%	13,0%	13,7%	12,4%	11,6%	7,4%	8,2%

The differences are not so significant, although there is a slight decrease in the quality of results. It can be observed, comparing 5.52 and 5.18, that there are less occurrences of side quantiles, with an approximation to the center.

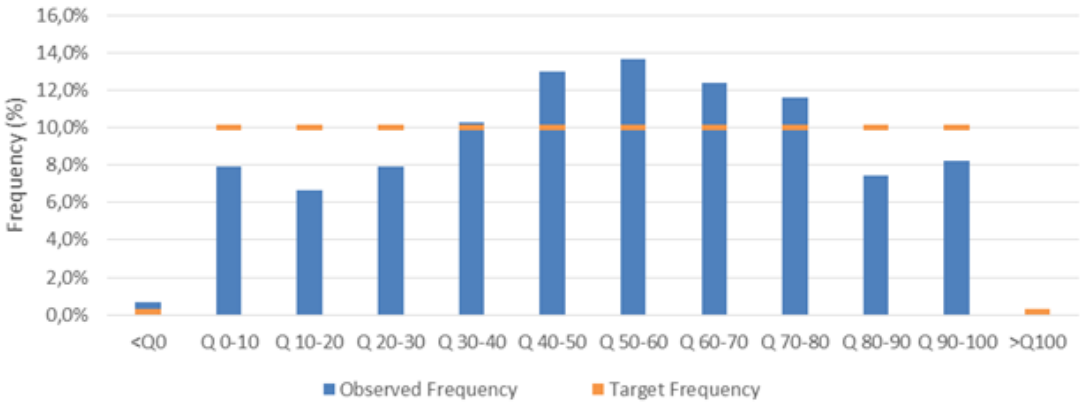

Figure 5.52: Quantile Distribution of B61 Voltage

The results shown are translated in a *RI* and *MAPE* of 77,2% and 0,83%, respectively, turning out to be a bit less accurate, overall.

Then, by selecting test cases as before, one for the winter test set and another for the summer test set, as the results are presented in tables 5.27 and 5.28.

Table 5.27: Beta Distributions Winter Set for B12 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	2,286982	3,258361
Beta	2,364291	3,494837
Min	61,46869	61,18121
Max	64,41211	64,82959

Table 5.28: Beta Distributions Summer Set for B12 Voltage

	0.5 Interval	Reactive Power Inputs
Alfa	5,461713	4,539248
Beta	5,621395	4,666376
Min	60,3609	60,18512
Max	65,80371	65,68835

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.53 and 5.54.

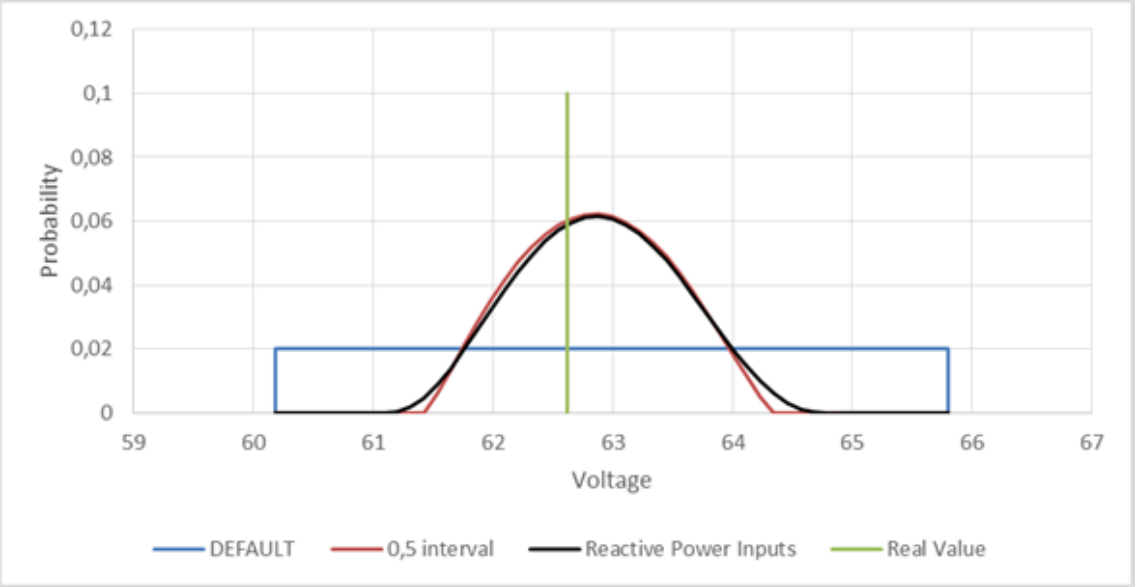

Figure 5.53: Beta Distribution of B12 Voltage, Winter Set

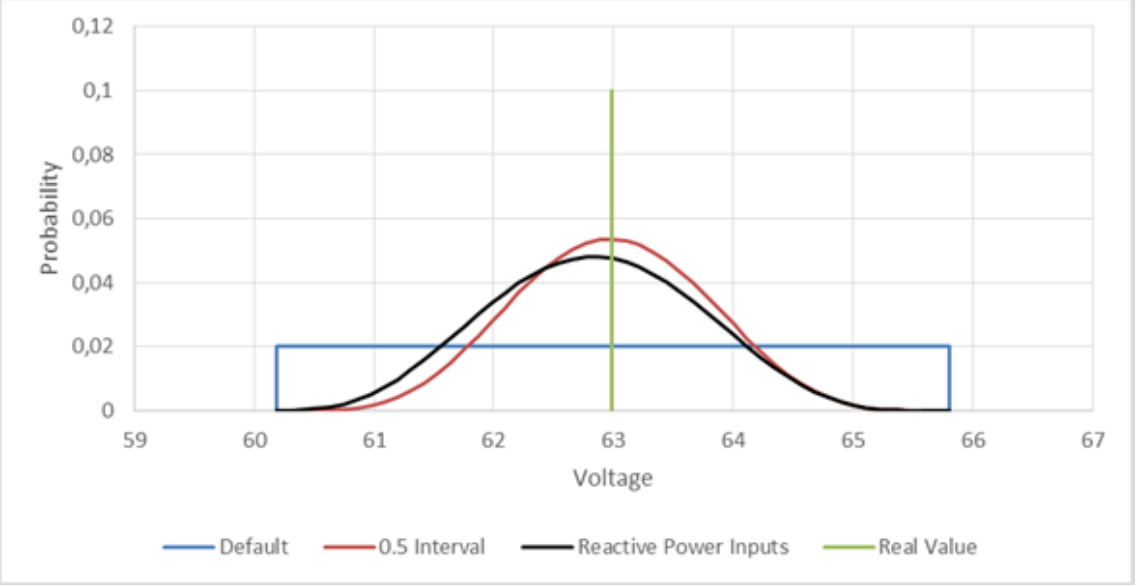

Figure 5.54: Beta Distribution of B12 Voltage, Summer Set

Again, even though the result comparison between the base case and this case tend to favour the base case, the figures present previously, 5.53 and 5.54, show that, for two different cases, that there is little difference between the results. In the winter set, both curves almost match each other, having the reactive power inputs case a wider interval, by almost 500 V in each limit.

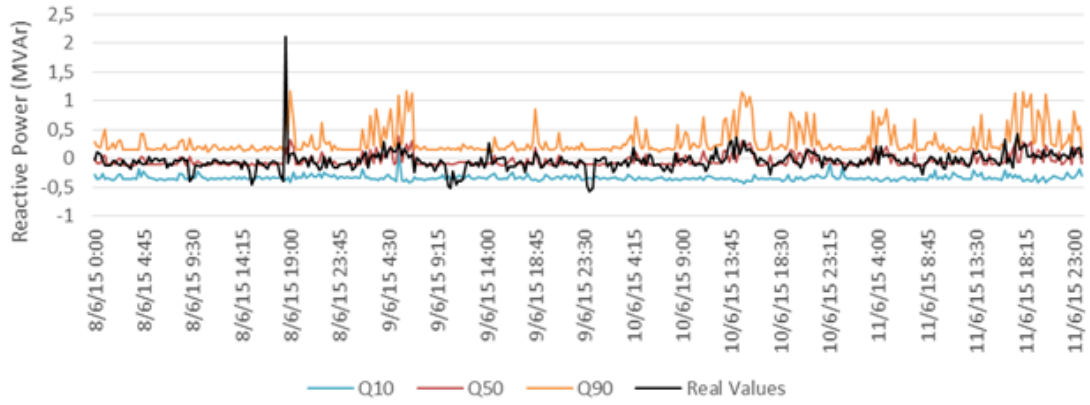

Figure 5.55: Quantile Distribution of B12 Reactive Power, Summer Set

However, for the summer set there is a slight tendency to shift the curve to the left, showing lower voltage values and assigning a lower probability to the real value. This can be again explained, just like B61, by the near zero values for the reactive power in B12 during summer set, presented in figure 5.55. This is another example where it is prejudicial to use reactive power inputs in the process of determination of voltage outputs.

5.4 Inclusion of Voltage Values

Finally, in this section will be evaluated, through the same means used in previous sections, the inclusion of voltage values in input data. This was saved for last mainly because of empirical reasons. Those reasons were that often voltage is not available to be measured directly, so it can not be a straight input. This happens because most of voltage reading devices such as transformers are located in the substations. The objective is then to understand the impact a known voltage input would have in the determination of other voltage outputs.

In order to do this, it was used a small portion of the grid, which only contemplates B6, B61, B62, which will have their active and reactive power as inputs and B4 as well, but accounting only for its voltage. In short, there will be three output variables (voltage of B6, B61 and B62) to be determined using ten input variables (B6, B61 and B62 active and reactive power, plus B4 voltage). It will be evaluated only on B6 and B61, for one main reason, that is both buses are of different kinds, one has active and reactive loads (B6) and the other has active and reactive power generation (B61), which would turn B62 evaluation obsolete. This case will be again compared to the base case described in section 5.1.

5.4.1 Bus 6 - Load Bus Connected to Generation

As expected, the difference is notorious between figures 5.56 and 5.1. The test set values obtained with voltage inputs are much more responsive to smaller changes of real values, which also indicates that there are less number of qualified cases found for each test case. This is due to a higher correlation between variables.

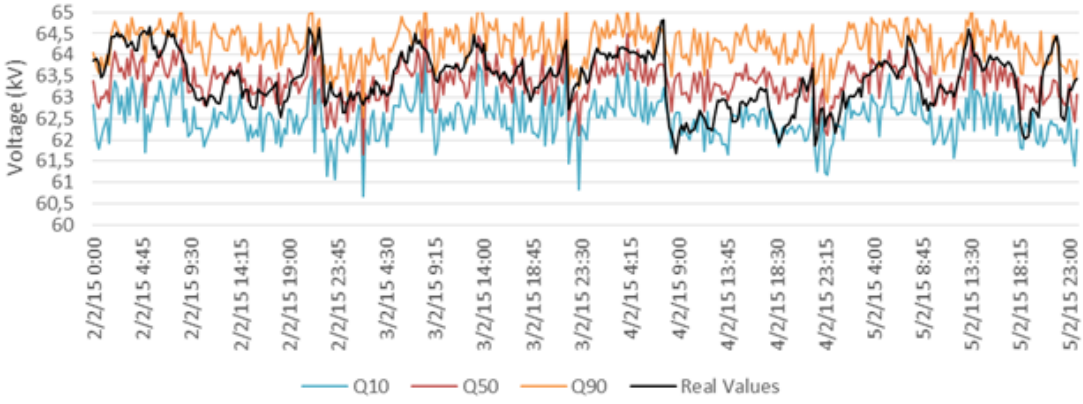

Figure 5.56: Winter Set B6 Voltage

Figures 5.57 and 5.58 show how the input is led by the real value, being translated in a good approximation for the winter set. For the summer set the input is not quite as good as the previous one.

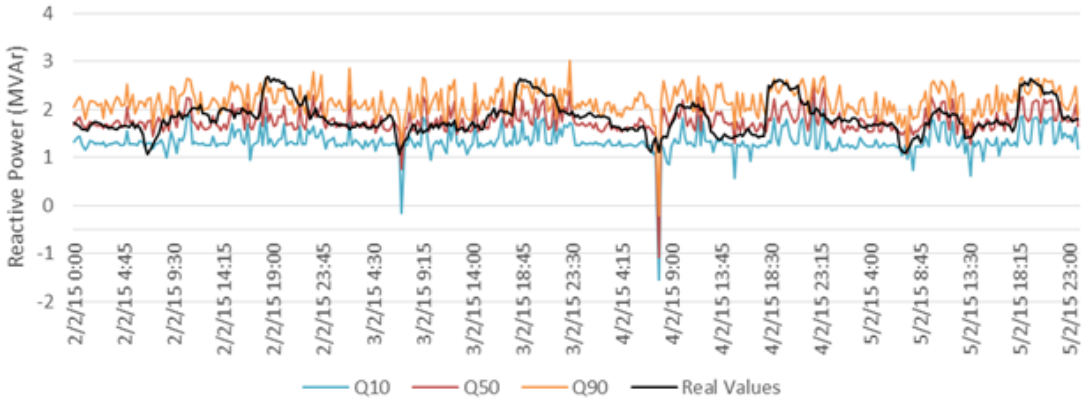

Figure 5.57: Quantile Distribution of B6 Reactive Power, Winter Set

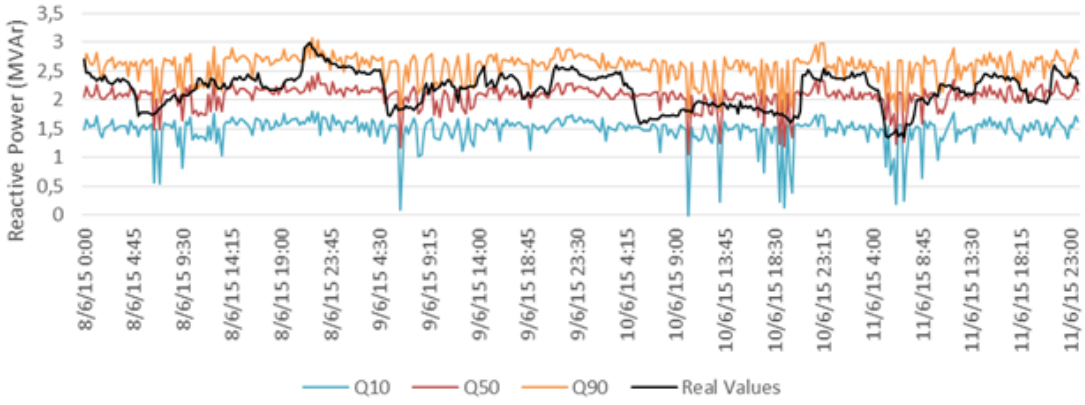

Figure 5.58: Quantile Distribution of B6 Reactive Power, Summer Set

By observing table 5.29, regarding quantile distribution, which is represented graphically in figure 5.60, one can draw some conclusions about the performance.

Table 5.29: Quantile Distribution of B6 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
7,8%	8,7%	10,3%	9,0%	11,7%	11,8%	10,9%	9,4%	9,2%	10,9%

With the information presented before, one can assess that there is a large improvement in this case by adding the only one voltage input. This is explained by the correlation between variables, presented next in figure 5.59.

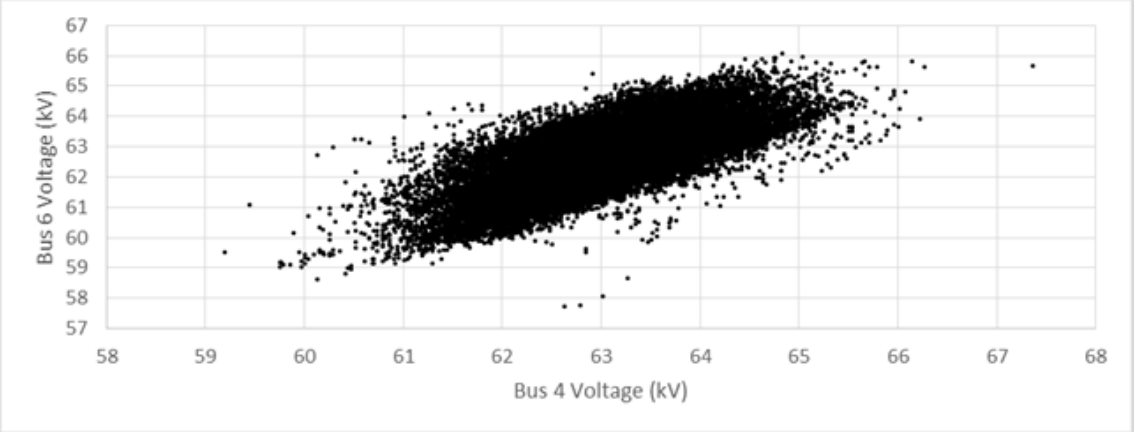

Figure 5.59: Correlation between B4 and B6 Voltages

As expected, figure 5.59 evidences a good correlation between B4 and B6 voltage. It is important to emphasize that the two buses are close to each other. A greater correlation could be expected between connected buses.

Given this and the information in table 5.29, one can assume the distribution spreads through the various quantiles equally.

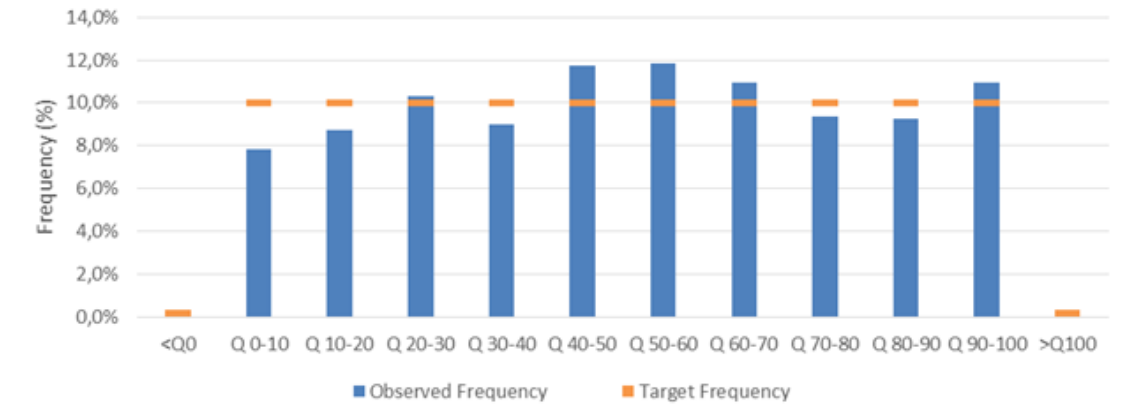

Figure 5.60: Quantile Distribution of B6 Voltage

The results shown are translated in a *RI* and *MAPE* of 88,4% and 0,90%, respectively, turning out to be more accurate than any of the previous cases. Another index was evaluated, just as for the base case, solely for B6 which contemplates the average difference between quantiles Q_{10} and Q_{90}. For this case the difference was of 1,88 kV. This difference, having 60 kV as per unit (p.u.) base voltage, represents an interval of 0.031 p.u., which deems itself acceptable.

The previous statement can be corroborated by a simple analysis of the distribution curves, from one case and another. In order to do that, a random test case was selected for the winter test set and another for the summer test set and the results are presented in tables 5.30 and 5.31.

Table 5.30: Beta Distributions Winter Set for B6 Voltage

	0.5 Interval	Voltage Input
Alfa	2,968623	3,475919
Beta	3,260226	3,555866
Min	61,29351	61,52808
Max	65,7176	65,91398

Table 5.31: Beta Distributions Summer Set for B6 Voltage

	0.5 Interval	Voltage Input
Alfa	6,585835	6,046204
Beta	6,567768	4,775207
Min	58,65869	58,6096
Max	65,81033	65,05207

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.61 and 5.62.

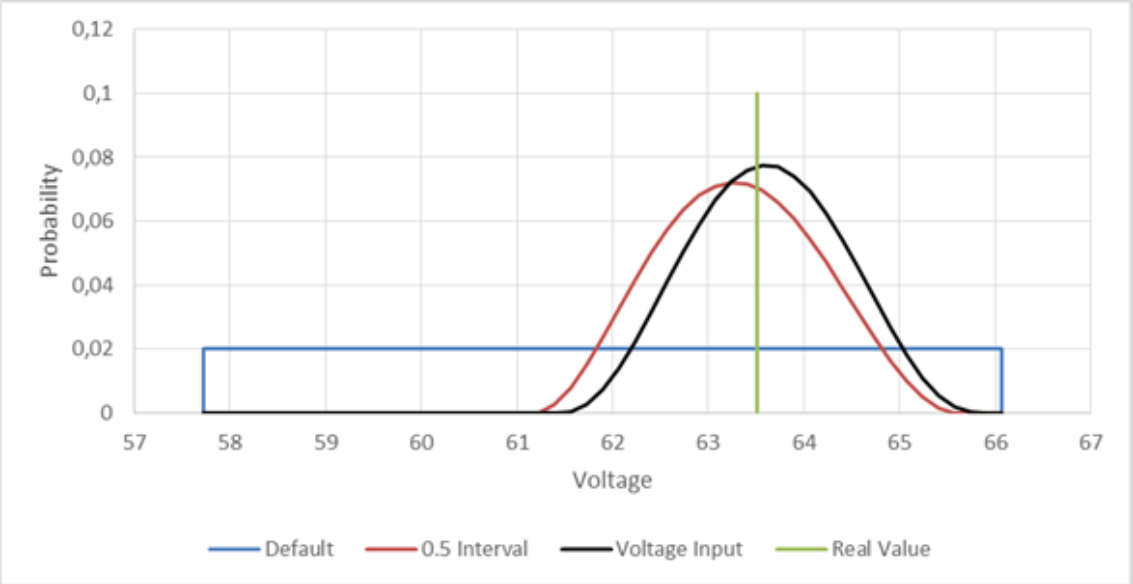

Figure 5.61: Beta Distribution of B6 Voltage, Winter Set

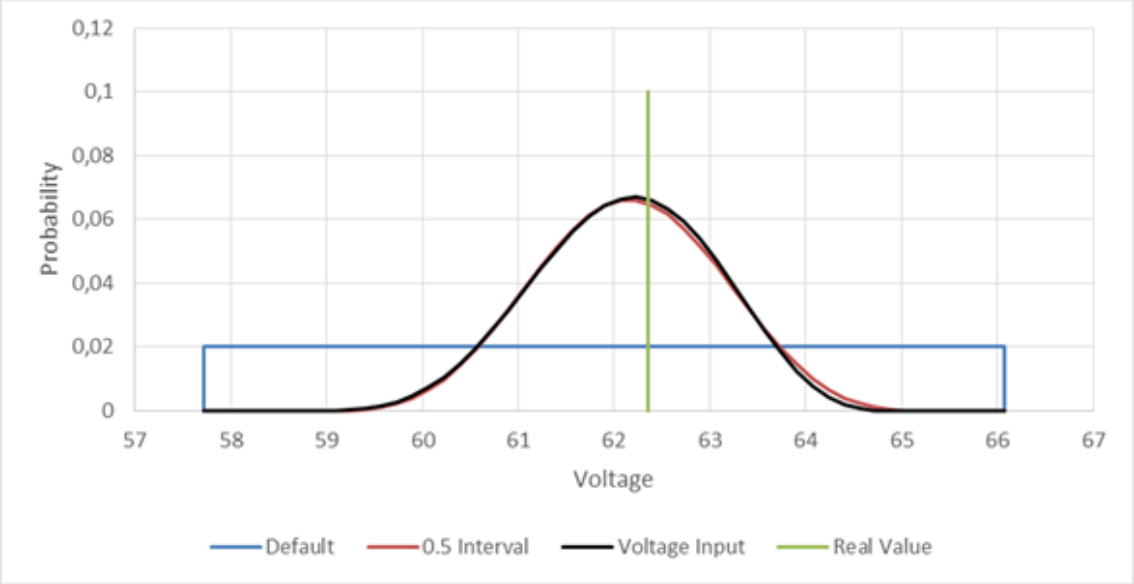

Figure 5.62: Beta Distribution of B6 Voltage, Summer Set

The results presented here are very important, since it involves an input, as previously stated, that is not always available.

Regarding the winter set, it is possible to observe that there are higher probabilities for the real value, as it is close to the maximum of the curve. One can observe also that there are higher values for the parameters α and β, which tightens the curve and assigns lower probabilities to the values closer to the limits.

By observing the summer set, one can observe the similarity of results between the base case and the one presented in this section. This is mainly due to the use of reactive power inputs, as it was studied before the possible nefarious effects of very low and not very distinctive values on voltages outputs. However, once again, it was determined that parameters α and β were higher in this case than in the base case, which could lead to better results if the interval is shrunk.

5.4.2 Bus 61 - Hydro Power Generation

Once again, and by far the easiest to assess, the difference is notorious between figures 5.63 and 5.11. The test set values obtained with voltage inputs are much more responsive to smaller changes of real values.

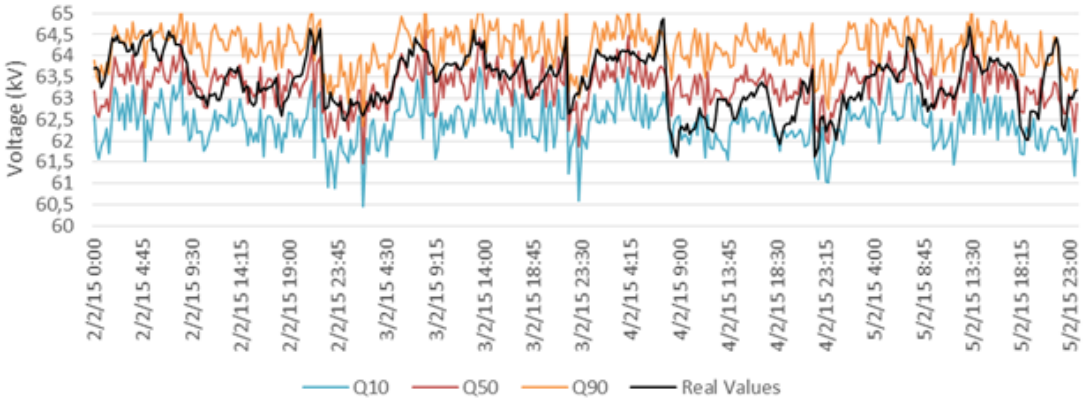

Figure 5.63: Winter Set B61 Voltage

Figures 5.64 and 5.65 show how the input is led by the real value, being translated in a good approximation for both winter set and summer set. One can observe the lack of diversity in the summer set, having a great number of values very close to zero.

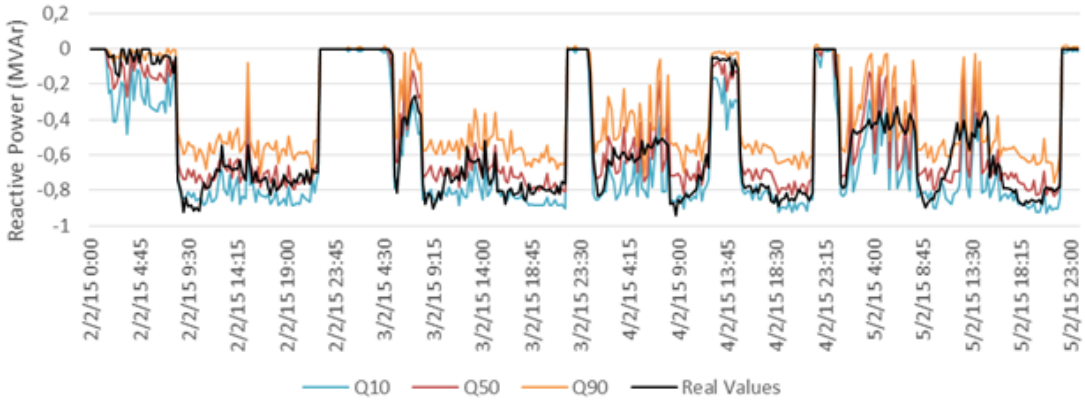

Figure 5.64: Quantile Distribution of B61 Reactive Power, Winter Set

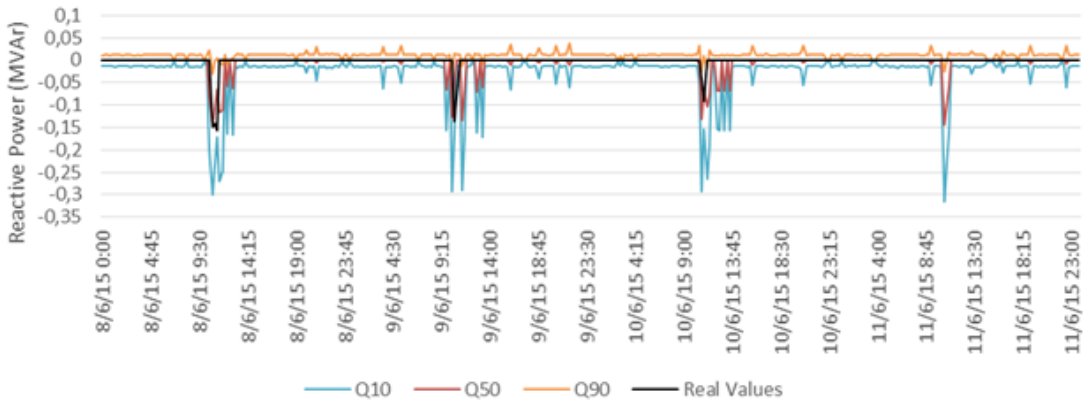

Figure 5.65: Quantile Distribution of B61 Reactive Power, Summer Set

Once again we resort to the quantile distribution, presented in table 5.32, being graphically explained in figure 5.67 to evaluate the performance.

Table 5.32: Quantile Distribution of B61 Voltage

Q_{0-10}	Q_{10-20}	Q_{20-30}	Q_{30-40}	Q_{40-50}	Q_{50-60}	Q_{60-70}	Q Q_{70-80}	Q_{80-90}	Q_{90-100}
7,0%	8,6%	8,6%	8,7%	11,6%	14,1%	9,5%	9,9%	10,5%	11,3%

With the information presented before, one can assess that there is a large improvement in this case by adding the only one voltage input. This is explained by the same reasons presented before, at B6, that is the correlation between variables, presented next in figure 5.66.

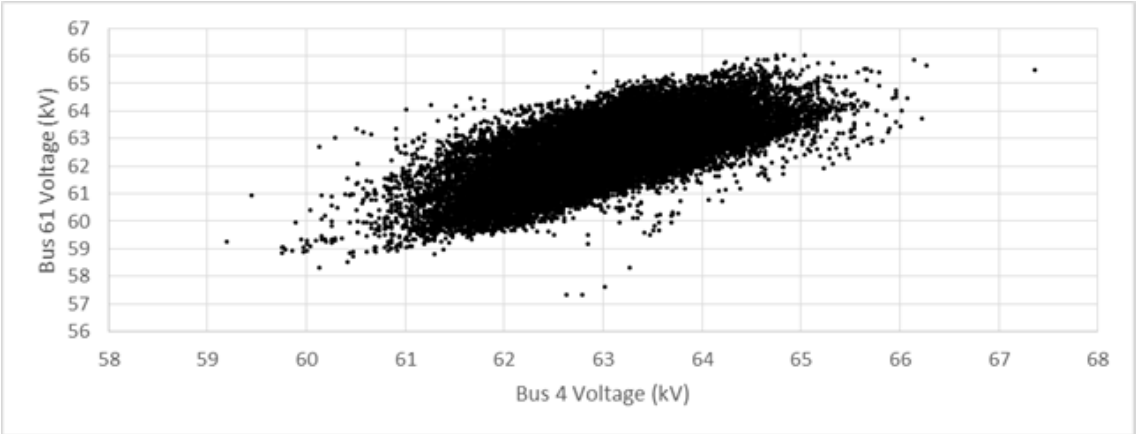

Figure 5.66: Correlation between B4 and B61 Voltages

As expected, figure 5.66 evidences a good correlation between B4 and B61 voltage. It is important to emphasize that the two buses, B4 and B61, are close to each other but not connected, since B61 is connected to B6.

Given this and the information in table 5.32, one can assume the distribution spreads through the various quantiles equally.

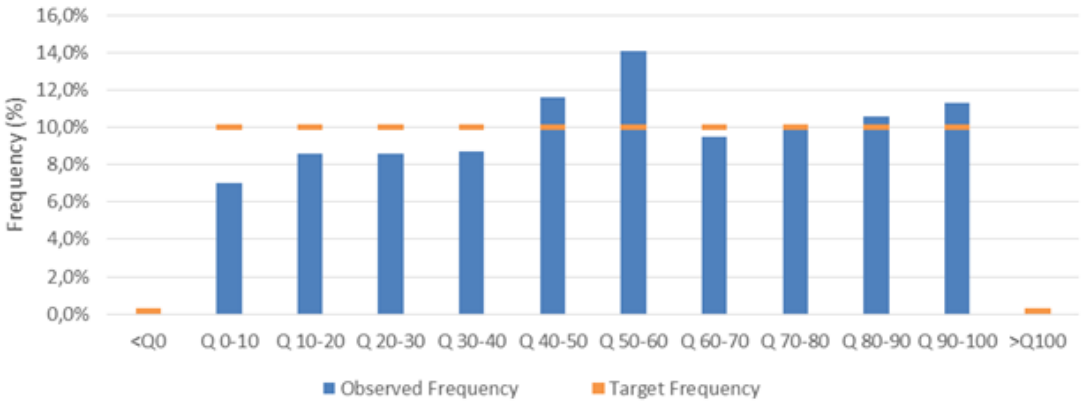

Figure 5.67: Quantile Distribution of B61 Voltage

The results shown are translated in a *RI* and *MAPE* of 84,8% and 0,93%, respectively, turning out to be more accurate than any of the cases.

The previous statement can be corroborated by a simple analysis of the distribution curves, from one case and another. In order to do that, the same test case was selected for the winter test set and for the summer test set and the results are presented in tables 5.33 and 5.34.

Table 5.33: Beta Distributions Winter Set for B61 Voltage

	0.5 Interval	Voltage Input
Alfa	2,983113	3,46539
Beta	3,33614	3,427965
Min	61,24232	61,40294
Max	65,7774	65,90966

Table 5.34: Beta Distributions Summer Set for B61 Voltage

	0.5 Interval	Voltage Input
Alfa	6,820653	6,267147
Beta	6,432024	4,975783
Min	58,32287	58,32287
Max	65,5223	64,93652

After retrieving this values, it is possible to construct the Beta PDF presented in the following figures 5.68 and 5.69.

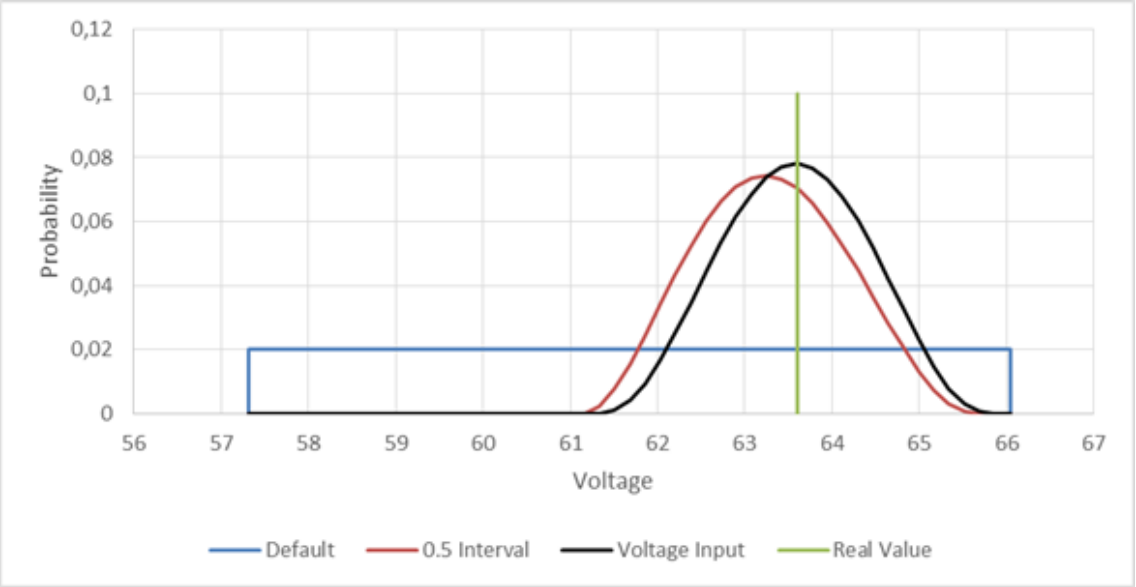

Figure 5.68: Beta Distribution of B61 Voltage, Winter Set

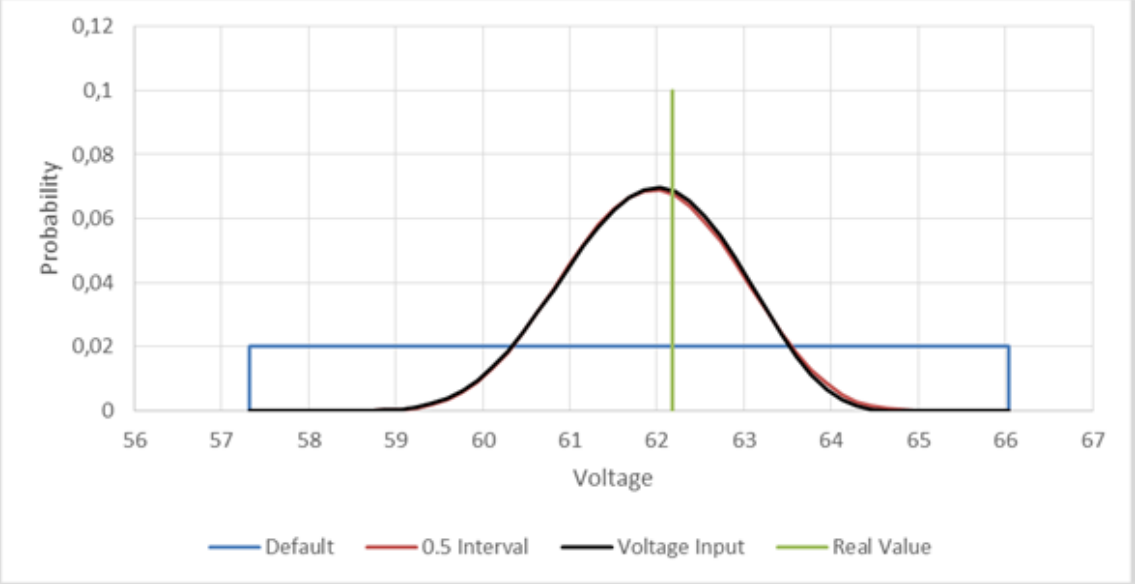

Figure 5.69: Beta Distribution of B61 Voltage, Summer Set

The results presented for this bus confirms what was previously written for B6. This results are important as there is a major improvement in using such information.

The analysis for this bus is in everything similar to the previous one, as the winter set revealed to be better than the base case, as the real value is very close to the maximum probability of the curve. In this case, the interval is almost the same, but the α and β parameters change slightly to make the curve tighter once again. The conclusions for the summer case are also the same, as it is observable the similarities between cases, almost matching both curves.

Chapter 6

Conclusions

This dissertation had the major objective to develop a probabilistic tool that would allow small modules throughout any information platform to perform a forecast study, based on historical information. The methodology was applied to an electrical grid as a means to solve a probabilistic load flow. However, the applications of the method presented in this thesis go far beyond the electrical utilities, being able to be also used on any forecast problem based on historical sets.

6.1 General Conclusions

Firstly, it was studied the current state of the art on PLF problems, to better understand what was at stake and what was the purpose of applying a probabilistic method. It was thoroughly described the main methods, the type of information used and on what kind of problems they were used. It was also done an overview of the current state of IoT applications and reactive programming paradigms.

Secondly, it was developed the mathematical tools to be applied. This was a major step, since all the mathematical approaches and program coding were developed already thinking in a wider application than electrical grids. This accounted for several changes throughout the course of the dissertation to be able to come up with an application that suits all kind of probabilistic problems that resort to historical data.

Thirdly, the information kindly given by EDP was studied. At first there were some problems regarding the information since it was not synchronized and had multiple values supposedly for the same instant. A correction was made to those values before synchronizing everything, which would later constitute the matrix of knowledge. At this point, the totality of the registries where separated in two sets: training set and test set. A further division was made to the test set into two groups, winter set and summer set.

Finally, the data was applied to test out the method presented. This step consisted on the evaluation of the influence of different components, such as the input interval amplitude, the use of a greater number of inputs acknowledging the reactive power and the incorporation of voltage inputs.

The first conclusion is that most of the results obtained were expected, as it made sense mathematically. On the base case, the results obtained set a high standard, as it was presented an average of 81,4% for the RI, 0,94% for the MAPE and 2,27 kV for the average difference between quantiles Q_{10} and Q_{90} on B6. The results obtained via shrank interval did not come as surprise, since the results went accordingly, presenting better distributions on selected cases which were confirmed by the performance indicators, just about 82,1% and 0,9375% for RI and MAPE, respectively, and 2,18 kV for the average difference between quantiles Q_{10} and Q_{90} on B6. This may not sound like a major improvement, but by evaluating the outputs individually, all buses, except for B12, had better RI values by almost 1% and the same for MAPE, but with exception of B7, presented a decreased value of almost 0,01%.

The next case presented the introduction of reactive power variables. This was expected to have some positive impact in the outputs, since the tests are being done to a high voltage grid and the inductance plays a major role on the definition of the voltage. However, the results are a lot worse than the base case, showing some biased results when evaluating the test case as a whole or as two separate groups. As a whole, the RI and MAPE average values dropped to 75,325% and 0,96%, respectively, which turned out to be significantly worse than the base case. Although these indexes point to worse applications, it was also very intriguing to evaluate the values of the average difference between quantiles Q_{10} and Q_{90}, since it topped the base case by a mere 0,01 kV, to 2,17 kV on B6. However, this is not enough to contradict the other indexes of RI and MAPE.

Evaluating the two groups individually, one can easily identify the problem. The fact that there was little production of reactive power from hydro power plants or wind farms during the summer actually worsened the input dataset, which in turn made the output qualified cases not to be the corresponding to the actual real value. During the winter, the results were satisfactory, as the distributions for all four buses, which the analysis focused on, were significantly better than the base case.

The last case was the most intriguing one, as the results where very much expected to be the best among the four cases, which came down as true. This case used the same interval as the base case and instead of using the active and reactive power of all buses as inputs, only B6, B61 and B62 had their active and reactive powers as inputs, being added the voltage on B4. The results were by far the best, averaging 86,6% and 0,915% of RI and MAPE, respectively, and 1,88 kV for the average difference between quantiles Q_{10} and Q_{90} on B6. The results were confirmed by the distributions presented and were justified by the correlation shown between the voltage on B4 and the rest of the output buses.

It can be concluded that this tool, besides being innovative, can also be useful in a panoply of problems. It can be used by a wind farm owner to forecast the production of his generation, through the use of meteorological data. The great thing about this new method is that it does not need necessarily the same types of data, that is, in order to forecast for voltages, it can be used all kinds of information, from chronological to electric, being the results solely as good as the correlation between input and output variables.

6.2 Objectives Satisfaction

The main objective of this dissertation focused on the development of a tool to solve PLF problems, with power to react to any change in state of the variables, as well as their synchronization. The tests were done firstly to a small part of the grid, to test and debug, and then applied to the real case example.

As referenced in the initial objectives, after the development of the described methodology, it was tested different cases, of a real case scenario, to asses the influence of different variables on the output results. The accuracy of results were based on three major indexes such as the Reliability Index (RI), the Mean Absolute Percentage Error ($MAPE$) and the average difference between quantiles Q_{10} and Q_{90} on one bus.

References

[1] N. Gershenfeld, R. Krikorian, and D. Cohen, *The Internet of things.*, 2004, vol. 291, no. 4. [Online]. Available: http://www.cba.mit.edu/docs/papers/04.10.i0.pdf

[2] E. Bainomugisha, A. L. Carreton, T. van Cutsem, S. Mostinckx, and W. de Meuter, "A survey on reactive programming," *ACM Computing Surveys*, vol. 45, no. 4, pp. 1–34, aug 2013. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2501654.2501666

[3] J. J. Grainger, W. D. Stevenson, and W. D. Stevenson, *Power system analysis*.

[4] G. J. Anders, *Probability concepts in electric power systems*. Wiley, 1990.

[5] R. Allan, C. Grigg, and M. Al-Shakarchi, "Numerical techniques in probabilistic load flow problems," *International Journal for Numerical Methods in Engineering*, vol. 10, no. 4, pp. 853–860, 1976.

[6] R. Allan and A. L. Da Silva, "Probabilistic load flow using multilinearisations," in *IEE Proceedings C (Generation, Transmission and Distribution)*, vol. 128, no. 5. IET, 1981, pp. 280–287.

[7] T. Sharaf and G. Berg, "Stochastic and probabilistic load flow analysis in system planning," *Canadian Electrical Engineering Journal*, vol. 8, no. 1, pp. 9–17, 1983.

[8] M. Brucoli, F. Torelli, and R. Napoli, "Quadratic probabilistic load flow with linearly modelled dispatch," *International Journal of Electrical Power & Energy Systems*, vol. 7, no. 3, pp. 138–146, jul 1985. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0142061585900420

[9] H. Wu, Y. Zhou, S. Dong, and Y. Song, "Probabilistic Load Flow Based on Generalized Polynomial Chaos," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 1–2, jan 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7456346/

[10] C. Delgado and J. Domínguez-Navarro, "Point estimate method for probabilistic load flow of an unbalanced power distribution system with correlated wind and solar sources," *International Journal of Electrical Power and Energy Systems*, vol. 61, pp. 267–278, 2014.

[11] "Basic Concepts of Correlation." [Online]. Available: http://www.real-statistics.com/correlation/basic-concepts-correlation/

[12] C. J. Roy and W. L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," *Computer Methods in Applied Mechanics and Engineering*, vol. 200, no. 25-28, pp. 2131–2144, 2011.

[13] J. Liu, X. Hao, P. Cheng, W. Fang, and S. Niu, "A Parallel Probabilistic Load Flow Method Considering Nodal Correlations," *Energies*, vol. 9, no. 12, p. 1041, dec 2016. [Online]. Available: http://www.mdpi.com/1996-1073/9/12/1041

[14] X. Li, J. Cao, and D. Du, "Probabilistic optimal power flow for power systems considering wind uncertainty and load correlation," *Neurocomputing*, vol. 148, pp. 240–247, 2015.

[15] M. Aien, M., Rashidinejad, M., Fotuhi-Firuzabad, "On Possibilistic and Probabilistic Uncertainty Assessment of Power Flow Problem: A Review and a New Approach," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 883–895, 2014.

[16] X. Fu, H. Sun, Q. Guo, Z. Pan, W. Xiong, and L. Wang, "Uncertainty analysis of an integrated energy system based on information theory," *Energy*, vol. 122, pp. 649–662, 2017.

[17] N. Soleimanpour and M. Mohammadi, "Probabilistic load flow by using nonparametric density estimators," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3747–3755, nov 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6514686/

[18] M. Aien, M. Fotuhi-Firuzabad, and F. Aminifar, "Probabilistic load flow in correlated uncertain environment using unscented transformation," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2233–2241, nov 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6192342/

[19] M. A. Abdullah, A. P. Agalgaonkar, and K. M. Muttaqi, "Probabilistic load flow incorporating correlation between time-varying electricity demand and renewable power generation," *Renewable Energy*, vol. 55, pp. 532–543, 2013.

[20] C. Chen, W. Wu, B. Zhang, and H. Sun, "Correlated probabilistic load flow using a point estimate method with Nataf transformation," *International Journal of Electrical Power and Energy Systems*, vol. 65, pp. 325–333, feb 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0142061514006383

[21] M. Pourahmadi-Nakhli, A. R. Seifi, and R. Taghavi, "A nonlinear-hybrid fuzzy/probabilistic load flow for radial distribution systems," *International Journal of Electrical Power and Energy Systems*, vol. 47, no. 1, pp. 69–77, 2013.

[22] P. Chen, Z. Chen, and B. Bak-Jensen, "Probabilistic load flow: A review," in *3rd International Conference on Deregulation and Restructuring and Power Technologies, DRPT 2008*. IEEE, apr 2008, pp. 1586–1591. [Online]. Available: http://ieeexplore.ieee.org/document/4523658/

[23] M. Kabir, Y. Mishra, and R. Bansal, "Probabilistic load flow for distribution systems with uncertain PV generation," *Applied Energy*, vol. 163, pp. 343–351, 2016.

[24] A. Jain, S. Tripathy, R. Balasubramanian, and Y. Kawazoe, "Stochastic load flow analysis using artificial neural networks," in *2006 IEEE Power Engineering Society General Meeting*. IEEE, 2006, p. 6 pp. [Online]. Available: http://ieeexplore.ieee.org/document/1709368/

[25] E. Zio, M. Delfanti, L. Giorgi, V. Olivieri, and G. Sansavini, "Monte Carlo simulation-based probabilistic assessment of DG penetration in medium voltage distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 852–860, jan 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0142061514005304

[26] H. Hong, "An efficient point estimate method for probabilistic analysis," *Reliability Engineering & System Safety*, vol. 59, no. 3, pp. 261–267, mar 1998. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0951832097000719

[27] M. Mohammadi, A. Shayegani, and H. Adaminejad, "A new approach of point estimate method for probabilistic load flow," *International Journal of Electrical Power & Energy Systems*, vol. 51, pp. 54–60, oct 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0142061513000793

[28] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for photovoltaic distributed generation using the Cornish-Fisher expansion," *Electric Power Systems Research*, vol. 89, pp. 129–138, aug 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378779612000776

[29] D. Villanueva, A. E. Feijóo, and J. L. Pazos, "An analytical method to solve the probabilistic load flow considering load demand correlation using the DC load flow," *Electric Power Systems Research*, vol. 110, pp. 1–8, may 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378779614000078

[30] Y. Liu, S. Gao, H. Cui, and L. Yu, "Probabilistic load flow considering correlations of input variables following arbitrary distributions," *Electric Power Systems Research*, vol. 140, pp. 354–362, nov 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378779616302139

[31] T. Institution of Electrical Engineers., *IEE proceedings. Generation, transmission and distribution.* [Institution of Electrical Engineers], 1994, vol. 142, no. 1. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/ip-gtd{_}19951484

[32] L. Pereira, V. da Costa, and A. Rosa, "Interval arithmetic in current injection power flow analysis," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 1106–1113, dec 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0142061512002244

[33] W. Briceno Vicente, R. Caire, and N. Hadjsaid, "Probabilistic load flow for voltage assessment in radial systems with wind power," *International Journal of Electrical Power & Energy Systems*, vol. 41, no. 1, pp. 27–33, oct 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0142061512000415

[34] Y. Liu, S. Gao, H. Cui, and L. Yu, "Probabilistic load flow considering correlations of input variables following arbitrary distributions," *Electric Power Systems Research*, vol. 140, pp. 354–362, nov 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0378779616302139

[35] J. Cao and Z. Yan, "Probabilistic optimal power flow considering dependences of wind speed among wind farms by pair-copula method," *International Journal of Electrical Power and Energy Systems*, vol. 84, pp. 296–307, 2017.

[36] A. L. Da Silva, J. M. Lima, S. Ribeiro, V. Arienti, M. T. Schilling, X. Vieira Filho, and S. Soares, "Operational and expansion planning of brazilian systems based on probabilistic load flow," in *Probabilistic Methods Applied to Electric Power Systems, 1991., Third International Conference on*. IET, 1991, pp. 197–202.

[37] R. Allan, C. Grigg, D. Newey, and R. Simmons, "Probabilistic power-flow techniques extended and applied to operational decision making," in *Proceedings of the Institution of Electrical Engineers*, vol. 123, no. 12. IET, 1976, pp. 1317–1324.

[38] A. L. Da Silva, S. P. Ribeiro, V. Arienti, R. Allan, and M. Do Coutto Filho, "Probabilistic load flow techniques applied to power system expansion planning," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1047–1053, 1990.

[39] K. Ashton, "In the real world, things matter more than ideas." [Online]. Available: http://www.rfidjournal.com/articles/pdf?4986

[40] K. Rose, S. Eldridge, and L. Chapin, "The Internet of Things: An Overview Understanding the Issues and Challenges of a More Connected World," 2015. [Online]. Available: https://pdfs.semanticscholar.org/6d12/bda69e8fcbbf1e9a10471b54e57b15cb07f6.pdf

[41] O. Vermesan, P. m. Friess, and River Publishers., *Internet of things: global technological and societal trends*. River Publishers, 2011.

[42] P. C. Evans and M. Annunziata, "Industrial Internet: Pushing the Boundaries of Minds and Machines," 2012. [Online]. Available: http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf

[43] G. Berry, "Real time programming: special purpose or general purpose languages, rapport de recherche inria, n 1065," 1989.

[44] E. A. Nadaraya, "On Estimating Regression," *Theory of Probability & Its Applications*, vol. 9, no. 1, pp. 141–142, jan 1964.

[45] G. S. Watson, "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 359–372, 1964.

[46] "Chuva e vento poupam 1,2 mil milhões a Portugal." [Online]. Available: https://www.dinheirovivo.pt/empresas/chuva-e-vento-poupam-12-mil-milhoes-a-portugal/

[47] "Evolução da potência instalada em renováveis por tecnologia [MW] - Apren | Associação de Energias Renováveis." [Online]. Available: http://www.apren.pt/fotos/editor2/graficos_com_titulo_17.jpg